

国立国語研究所学術情報リポジトリ

均衡会話コーパス設計のための一日の会話行動に関する基礎調査

メタデータ	言語: Japanese 出版者: 公開日: 2016-01-29 キーワード (Ja): キーワード (En): survey of conversational behavior, corpus design, corpus of conversational Japanese 作成者: 小磯, 花絵, 土屋, 智行, 渡部, 涼子, 横森, 大輔, 相澤, 正夫, 伝, 康晴, KOISO, Hanae, TSUCHIYA, Tomoyuki, WATANABE, Ryoko, YOKOMORI, Daisuke, AIZAWA, Masao, DEN, Yasuharu メールアドレス: 所属:
URL	https://doi.org/10.15084/00000810

均衡会話コーパス設計のための一日の会話行動に関する基礎調査

小磯花絵^a 土屋智行^b 渡部涼子^c 横森大輔^b 相澤正夫^d 伝 康晴^e

^a 国立国語研究所 理論・構造研究系

^b 九州大学

^c 国立国語研究所 コーパス開発センター 非常勤研究員

^d 国立国語研究所 時空間変異研究系

^e 千葉大学／国立国語研究所 言語資源研究系 客員教授

要旨

本稿では、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」(リーダー:小磯, 2014年7月～2015年8月)の活動について報告する。本プロジェクトの目標は、21世紀初頭の日本語母語話者の多様な会話行動を納めた大規模な日本語日常会話コーパスの構築を目指す、その基盤整備として、(1)均衡性を考慮した会話コーパスの設計、(2)種々の日常場面での会話を収録するための方法論、(3)日常会話を適切・効率的に転記するための方法論の策定を進めることである。本稿ではこのうち(1)に焦点を当て、均衡性を考慮したコーパス設計案を策定するために実施した、一日の会話行動の種類と従事時間に関する調査について報告する。調査では、首都圏在住の成人約250人を対象に、起床から就寝までの間に行ったそれぞれの会話について、いつ、どこで、誰と、何をしながら、どのような種類の会話を、どのくらいの長さ行ったか、などを問う調査項目に回答してもらった。その結果、日常会話には以下の傾向が見られることが分かった。①雑談や用談・相談が多く、会議・会合や授業・レッスン・講演は少ない。②少人数・短時間の会話が多い。③自宅や職場・学校など、私的あるいは公的に主たる空間での会話が多い。④いくつかの調査項目の間には対応関係が見られる。これらの分析結果にもとづき、均衡性を考慮した日常会話コーパスの設計方針について議論する*。

キーワード: 会話行動調査, コーパス設計, 日常会話コーパス

1. はじめに

日常会話は社会生活の基盤であり、日常の話し言葉の特徴や仕組み、日常生活を円滑にするための会話コミュニケーションの有様を解明することが求められている。こうした研究を支えるものとして、実際の日常会話場面を対象とした大規模な会話コーパスの構築が不可欠である。また、言葉や行動様式は常に変化しているため、こうしたコーパスは、後世の人々が21世紀初頭の日本語母語話者の言語生活を知るための貴重な記録となる。民俗文化的価値のある日常会話を記録・保存し、後世に伝えることは、現代を生きる我々に課された重要な課題である。

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)では、21世紀

* 本稿は国立国語研究所フィージビリティ型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)の研究結果である。また、本稿の内容は、2015年2月24日に国立国語研究所で開催された第123回NINJALサロンでの口頭発表および2015年3月10日に開催された第7回コーパス日本語学ワークショップでの口頭発表(小磯ほか2015)の内容を加筆・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下された方々に感謝申し上げる。

初頭の日本語母語話者の多様な会話行動を納めた大規模な日本語日常会話コーパスの構築を目指し、その基盤整備として、以下の3種の検討を進めている。

- ① 均衡性（バランス）を考慮した会話コーパスの設計¹
- ② 種々の日常場面での会話を収録するための方法論
- ③ 日常会話を適切・効率的に転記するための方法論

言語生活の本質を解明するには、日常の言語生活を反映したコーパス設計が不可欠である。また、屋外での会話や携帯電話を介した会話など、種々の日常場面での会話を収録するための技術的・倫理的な問題を整理・解決し、日常会話を収録するための方法論を具体的に策定する必要がある。更には、大量に収録した会話を適切かつ効率的に転記するための方策を確定することが求められる。

本研究プロジェクトは、コーパス構築に欠かすことのできない上記3点を具体的に検討・策定することで、大規模日常会話コーパス構築のための道筋をつけることを目指すものである。

表1に日本語の日常会話を納めた主要なコーパス・データベースの一覧を示す。個々のコーパスを見ると、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。また全体として見ても、親近者同士、中でも若者の雑談が多い。収録のし易さが一因と考えられるが、日常会話は親近者同士の雑談だけで構成されるわけではない。

表1 主要な日本語日常会話コーパス

コーパス名	規模	概要	音声
名大会話コーパス	161名 100時間	親しい者同士の雑談	無
BTSによる多言語話し言葉コーパス（宇佐美2007）	294会話 66時間	友人同士の雑談、教師学生面談会話、電話会話など	一部
Sakura コーパス	18会話	大学生の会話	有
千葉大学3人会話コーパス（伝・榎本2014）	12会話 2時間	大学生の友人同士の会話	有
CallHome Japanese（伝・フライ2000）	120会話 20時間	アメリカ在住日本人と国内の家族・友人との電話会話	有
CallFriend Japanese	31会話	アメリカ在住の日本人同士の電話会話	有
女性のことば・職場編（現代日本語研究会1999） 男性のことば・職場編（現代日本語研究会2002）	各21名	職場のフォーマル・インフォーマルな場面の自然談話	無

我々の言語生活を正確に記述し、その本質を解明するためには、日常の言語生活の幅広いレジスターをカバーするようサンプルを選定することが求められる。『現代日本語書き言葉均衡コーパス』では、書き言葉の生産、流通、受容の各過程が書き言葉の実態をとらえる上で重要とした上で、出版データと図書館収蔵図書を母集団としたランダムサンプリングを行い、生産実態と流

¹ コーパス設計において「均衡性」とは、対象言語（この場合は日本語日常会話）の全体像を把握するために、対象言語の幅広いレジスターをカバーするようサンプルを選定することを意味する。類似した用語として「代表性」があるが、これは対象言語の母集団の縮図となるようサンプルを選定することを意味する。

通実態を反映したサブコーパスを設計した (Maekawa et al. 2014)。しかし話し言葉の場合、多様なレジスターが存在することは十分予想されることだが、実際にどのようなレジスターが存在するか把握することはそれ自体が重要な研究課題である。

そこで本プロジェクトでは、日本語母語話者が日常的に交わす会話の生産実態をとらえてコーパス設計に活かすために、我々が普段どのような種類の会話をどの程度行っているかを調査した。本稿では調査の結果について報告する。

本稿の構成は次の通りである。まず2節において調査の概要について説明し、3節に調査結果を示す。また4節では、本調査に基づき日常会話コーパスをどのように設計するか、その方向性について議論する。

2. 調査の概要

2.1 調査項目の設計

会話行動を調査するにあたっては、我々が普段行っている会話をどのような視点でとらえるかが問題となる。本調査では、会話行動を大きく次の三つの軸でとらえることとし、具体的に、以下の表2から表4に示す調査項目を設定した。

- ① 会話の属性：会話の形式や会話の長さなど
- ② 会話状況の属性：会話の行われた時間帯や場所、活動など
- ③ 調査協力者（以下、協力者）の属性：性別や年代、職業など

以下ではこのうち、いくつかの調査項目について補足する。

表2 会話の属性に関する調査項目

項目	説明	回答方式	選択肢
形式	会話のタイプ	単一選択式	雑談, 用談・相談, 会議・会合, 授業・レッスン・講演
長さ	会話の長さ	単一選択式	5分未満, 5～15分, 15～30分, 30分～1時間, 1～2時間, 2～5時間, 5～10時間, 10時間以上
相手関係と人数	会話相手との関係	選択式（複数可）	家族, 親戚, 先生生徒, 仕事学業関係, 友人知人, 公共商業関係, 顔見知り・見知らぬ人
モード	関係ごとの人数	数値入力式	
相手属性	電話・ネットでの音声・映像会話	オプション式	（該当する場合に選択）
言語	外国人を含む会話	オプション式	（該当する場合に選択）
	外国語での・外国語を含む会話	オプション式	（該当する場合に選択）

表3 会話状況の属性に関する調査項目

項目	説明	回答方式	選択肢
時間帯	会話が行われた時間帯	単一選択式	午前, 午後, 夜
場所	会話が行われた場所	単一選択式	自宅, 職場・学校, 公共商業施設, 交通機関, それ以外の屋内, それ以外の屋外
活動	会話中にしていた活動	単一選択式	食事, 家事・雑事, 身の回りの用事, 療養, 仕事・学業, 業務外・課外活動, 社会参加, レジャー活動, 付き合い, 移動, 休息

表 4 協力者の属性に関する調査項目

項目	説明	回答方式	選択肢
性別	協力者の性別	単一選択式	男性, 女性
年代	協力者の年代	単一選択式	20 代, 30 代, 40 代, 50 代, 60 代以上
職業	協力者の職業	単一選択式	会社員・役員・公務員・専門職 (以下, 会社員等), 自営業, パート・アルバイト, 学生, 専業主婦, 無職・定年退職者 (以下, 無職等), その他
世帯員数	協力者の世帯員数	数値入力式	
居住地	協力者の居住地	単一選択式	東京都, 神奈川県, 千葉県, 埼玉県

■形式 表 2 に示した通り, 《形式》の選択肢として, 「雑談」「用談・相談」「会議・会合」「授業・レッスン・講演」を設けた。

「雑談」は, 会話の目的や話題などがあらかじめ定められていない会話を, 「用談・相談」は, 会話の目的はある程度決まっているが時間や場所などは定められていない会話を, 「会議・会合」は, 「用談・相談」とは異なり時間や場所などが定められている会話を, 「授業・レッスン・講演」は, 先生や講演者など会話の流れを導く人物がいる場での会話を指す。この選択肢の設定は, 国立国語研究所 (1971, 1987) および畠 (1983) で述べられている話し言葉の分類を参考にしている。

国立国語研究所 (1971) では, コミュニケーション上の機能にもとづき, 談話を「ひとり」「あいさつ」「しらせ・用談」「おしゃべり」「あそび」「教え・さしず」「けんか」「思考」に分類している。また畠 (1983) は, 計画性の程度にもとづき, 言語行動の場面を「拘束場面」と「自由場面」に分類している。「拘束場面」は, コミュニケーションの目的, 参加者の役割, 話題, コミュニケーションの手続きなどが外的に決まっている場面や, 言語行動が儀式化されている場面であるのに対し, 「自由場面」は, こうした場面や言語行動の様式があらかじめ決まっておらず, その場面の参加者が協力し自発的に決めていくような自由度の高い場面とされる。

本調査では, 以上の分類を参考に, 「おしゃべり」や「あそび」を自由場面としてまとめて「雑談」に, 目的をもつ「しらせ・用談」を拘束性の低い「用談・相談」と拘束性の高い「会議・会合」に分けた。また「拘束場面」の「教え・さしず」は「授業・レッスン・講演」とした。なお本調査では, あいさつのような極めて短いやりとりで構成される会話や独話は対象外とすることから, 「ひとり」「あいさつ」「思考」に相当する選択肢は設けていない。

■長さ 《長さ》は, 会話を含む場面の長さではなく, 実際に会話した時間に基づき判断することとした (具体的な選択肢は表 2 参照)。たとえば, 30 分の食事の中で会話をほとんどせず, その合計が 5 分未満であった場合, 30 分ではなく 5 分未満とする。

■相手関係と人数 国立国語研究所 (1980) の分類 (「家族関係」「師弟関係」「職場関係」「主・客関係」「知友関係」「その他」) を参考に, 表 2 に示す七つの選択肢を設定した。「公共商業関係」は, ものの売り買いやサービスなどの取引を行う関係を指し, 役所・店舗・飲食店などでの窓口・店員と客との間柄や, 病院での医師と患者との間柄などが該当する。「知人・友人」は, 互いをよく知っている, あるいは互いに名前と顔が一致する間柄を, 「顔見知り・見知らぬ人」はお互いの素性をよく知らない相手を指す。

■**場所** 国立国語研究所（1980）の分類を参考に、表3に示す六つの選択肢を設けた。国立国語研究所（1980）は、場所を大きく「公的生活」「私的生活」「外出先」に分類し、「外出先」をさらに「各種学校など」（語学学校など）、「アルバイト先」「公共施設」「店」「旅先」「路上」「駅」「車内」に分類している。

本調査では、「公的生活」を「職場・学校」、「私的生活」を「自宅」とした。また「外出先」のうち、公共商業的行為をいとなむ場所（「公共施設」「店」、遊園地など）を「公共商業施設」、移動をともしなう場所（「駅」「車内」）を「交通機関」と「それ以外の屋外」（「路上」や公園など）、これに「それ以外の屋内」（知人・親戚の家など）を加えて4項目に整理した。

■**活動** 《活動》の選択肢は、日本放送協会（2010）による国民生活時間調査の行動分類（中分類）を参考に表3に示す11の選択肢を設けた。

日本放送協会（2010）の中分類のうち、会話行動をともしなわない「睡眠」や他活動と共に起して現れる「マスメディア接触」は除外した。また「通勤」と「通学」は、家と店舗の往復などその他の移動と合わせて「移動」に、「仕事関連」と「学業」は「仕事・学業」にまとめた。仕事のつきあいや部活動など仕事や学業から派生する副次的な活動は、「業務外・課外活動」として新たな選択肢を設けた。その他の選択肢は日本放送協会（2010）の中分類に準拠した。ただし、選択肢の名称は分かりやすさを優先して一部変更した。「食事」は自宅・職場・店舗・屋外などでの食事が、「家事・雑事」は掃除や買物、子どもの世話などが、「身の回りの用事」は入浴や散髪などが、「療養」は通院や入院などが、「社会参加」は冠婚葬祭や町内会の行事などが、「レジャー活動」は趣味・娯楽・行楽・スポーツ・習いごとなどが、「付き合い」は知人との電話でのおしゃべりや同窓会など人と会うこと・話すことを主な目的とする活動が、「休息」は自宅での一家団らんや職場での休憩などが、それぞれ該当する。

2.2 調査の方法

本調査の目的は、多様な日常場面での会話を網羅したコーパスを設計するために、私たちが普段、どのような種類の会話をどの程度行っているかを明らかにすることである。構築を計画しているコーパスが首都圏で交わされる会話を対象とするため、調査の対象者も首都圏在住者に限定した。また、会話行動の実態の解明を目的とする場合、仮に丸一日会話をしない日があっても、それが生じる以上、調査の対象とすべきである。しかし本調査は、コーパス設計のために会話行動の多様性をとらえることを主目的としており、これを限られた期間と予算で達成するために、あまり会話しないと予想される日は調査日としないよう依頼した。その意味において、会話行動の正確な実態調査にはなっていない点に注意する必要がある。

調査方法の概要を以下に示す。

■**期間** 2014年11月1日～2015年2月28日

■**調査対象** 首都圏（東京・神奈川・千葉・埼玉）に在住の20歳以上の日本語母語話者243人に調査に協力してもらった。各世代（20代・30代・40代・50代・60代以上）の男女がそれぞれ

① どんな会話が			
<自由にメモしてください>			
自宅で夕食を食べながら家族と雑談			
② いつ (1つ選択)			
<input type="checkbox"/> 午前	<input type="checkbox"/> 午後	<input checked="" type="checkbox"/> 夜(午後6時頃～)	
③ どのくらい (1つ選択)			
<input type="checkbox"/> 5分未満	<input type="checkbox"/> 5～15分	<input type="checkbox"/> 15～30分	<input checked="" type="checkbox"/> 30分～1時間
<input type="checkbox"/> 1～2時間	<input type="checkbox"/> 2～5時間	<input type="checkbox"/> 5～10時間	<input type="checkbox"/> 10時間以上
④ どこで (1つ選択)			
<input checked="" type="checkbox"/> 自宅 <input type="checkbox"/> 職場・学校 <input type="checkbox"/> 公共商業施設 <input type="checkbox"/> 交通機関			
<input type="checkbox"/> それ以外の屋内 <input type="checkbox"/> それ以外の屋外			
⑤ だれと (あてはまるものをそれぞれに人数を記入)			
家族: 3 人 親戚: _____ 人 先生・生徒: _____ 人			
仕事・学業関係: _____ 人 公共商業関係: _____ 人			
友人・知人: _____ 人 顔見知り・見知らぬ人: _____ 人			
⑥ 何をしながら (1つ選択)			
<input checked="" type="checkbox"/> 食事 <input type="checkbox"/> 家事・雑事 <input type="checkbox"/> 身周りの用事 <input type="checkbox"/> 療養			
<input type="checkbox"/> 仕事・学業 <input type="checkbox"/> 業務外・課外活動 <input type="checkbox"/> 社会参加			
<input type="checkbox"/> レジャー活動 <input type="checkbox"/> 付き合い <input type="checkbox"/> 移動 <input type="checkbox"/> 休息			
⑦ どんな種類 (1つ選択)			
<input checked="" type="checkbox"/> 雑談 <input type="checkbox"/> 用談・相談 <input type="checkbox"/> 会議・会合 <input type="checkbox"/> 授業・レッスン・講演			
⑧ その他 (あてはまるものすべて選択)			
<input type="checkbox"/> 電話・スカイプなどの遠隔での音声・映像会話			
<input type="checkbox"/> 外国人を含む会話 <input type="checkbox"/> 外国語を含む会話			

図1 調査票のサンプル

れ20～25名となるよう調整した。調査協力者は、ホームページおよび知人などからの紹介により募集した。家庭での会話行動の多様性をとらえるため、家族等と同居している人を優先して募集し、各世帯1名を上限とした。

■調査日・時間 調査は、協力者1人あたり任意の平日2日・休日1日の計3日とし、起床してから就寝するまでの会話を記録するよう依頼した。

■調査項目 調査項目は、2.1節で述べた通り、会話に関する調査項目(表2の6項目)、会話状況に関する調査項目(表3の3項目)、協力者の属性(表4の5項目)とした。

このうち会話・会話状況に関する調査項目と会話の概要は「調査票」(図1)に、協力者の属性は「基本情報シート」にまとめた。

2.3 調査手続き

- 協力者に、調査の手引き、調査票(1日1冊、計3冊)、基本情報シート1部を事前に郵送した。
- 調査の手引きに目を通し、質問があれば電話あるいはメールで問い合わせるよう依頼した。
- 資料が届いてから2週間以内を目途に、協力者本人が任意の平日2日・休日1日(計3日)を選んで調査を実施してもらった。有職者は勤務日を平日、非勤務日を休日、それ以外は月～金を平日、土日祝日を休日とした。あまり会話しないと予想される日はできるだけ調査日としないよう依頼した。
- 調査日当日、協力者は調査票1冊(38頁分)を携帯し、起床してから就寝までの間に行った

全ての会話について、会話の概要を記した上で、会話と会話状況に関する調査項目(表2, 3)に回答してもらった。できるだけ一まとまりの会話が終了するごとに記録するよう依頼した。

- e) 調査終了後, 3 日分の会話行動を記録した調査票 3 冊と, 協力者の属性を記した基本情報シート 1 部を調査者に返送してもらった。
- f) 3 日間の調査協力に対し 6000 円の謝礼を支払った。

3. 調査結果

3.1 結果の概要

調査の応募者数は 266 名, うち有効な回答者数は 243 人 (729 日分, 3 日 /1 人) であった (有効回答率: 91.4%)。調査対象者 243 人の総会話数は 9,272 件, 1 日の平均会話数は 12.7 件 (平日: 13.4 件, 休日: 11.3 件), 1 日の平均会話時間は 6.2 時間 (平日: 6.1 時間, 休日: 6.5 時間) であった。なお, 会話時間は会話の長さの平均値 (例: 「1 ~ 2 時間」であれば 1.5 時間) から算出した。調査依頼時にあまり会話しないと予想される日は避けるよう依頼したため, 実態よりも 1 日の会話数や会話時間は多いことが予想される。

■協力者の属性 表 5 に, 協力者の属性のうち《性別》《年代》《職業》の内訳を示す。当初予定した通り, 各世代の男女につき 20 ~ 25 名を確保することができた。《職業》については, 男性は会社員等が半数を超えているのに対し, 女性は会社員等に加え, パート・アルバイトや専業主婦も多くを占めていた。男女ともに 20 代の大半は学生であった。《世帯員数》については, 1 人 (一人暮らし) が 32 名, 2 人が 62 名, 3 人が 61 名, 4 人が 50 名, 5 人以上が 14 名と, 家族と同居している割合が多かった。これは, 先述の通り, 家庭での会話行動の多様性をとらえるため家族等と同居している人を優先して募集したためである。

表 5 協力者の属性の内訳 (人 / 割合)

		20 代	30 代	40 代	50 代	60 代～	合計
男性	会社員・役員・公務員・専門職	5	20	23	18	7	73 (61.3%)
	自営業	0	0	2	2	4	8 (6.7%)
	パート・アルバイト	0	1	0	1	0	2 (1.7%)
	専業主婦	0	0	0	0	0	0 (0.0%)
	学生	18	0	0	0	0	18 (15.1%)
	無職・定年退職者	0	1	0	1	12	14 (11.8%)
	その他	0	0	0	2	2	4 (3.4%)
	小計	23	22	25	24	25	119
女性	会社員・役員・公務員・専門職	5	11	4	2	4	26 (21.0%)
	自営業	0	1	2	0	1	4 (3.2%)
	パート・アルバイト	1	5	8	12	4	30 (24.2%)
	専業主婦	1	5	8	8	11	33 (26.6%)
	学生	18	0	0	0	0	18 (14.5%)
	無職・定年退職者	0	0	1	1	2	4 (3.2%)
	その他	0	3	2	1	3	9 (7.3%)
	小計	25	25	25	24	25	124
	合計	48	47	50	48	50	243

■**会話の属性** 表6～11に、総会話数9,272件に関する会話の属性6項目の結果を示す。なお、表に記した数値は当該カテゴリー（選択肢）の出現件数と項目内での割合である。

《形式》については、雑談が全体の60%強を占めており、ついで用談・相談が30%強であった。一方、会議・会合や授業等はいずれも3%前後に留まった（表6）。このように日常会話では、雑談や用談・相談がその大半を占めていることが分かる。《長さ》は、15分以下の短い会話が約60%、30分以下の会話が約75%を占めており、1時間を超える会話は12%弱に留まる（表7）。《相手人数》については、全ての関係性の人数の合計値を用いた（表8）。表から、《相手人数》が1人・2人の場合（2～3人の会話）が全体の75%を占めており、日常会話では少人数の会話が多いことが分かる。《モード》については、対面会話が90%、電話やインターネットなどを介した遠隔会話は10%であり、遠隔会話も少なからず見られる（表9）。一方《相手属性》や《言語》の結果から、外国人を含む会話や外国語での、あるいは外国語を含む会話はいずれも全体の1%に満たず、一般的に日常会話ではあまり生じないケースと言える（表10,11）。

表6 形式（件／割合）²

雑談	5,719	(61.9%)
用談・相談	2,997	(32.4%)
会議・会合	317	(3.4%)
授業・レッスン・講演	209	(2.3%)

表7 長さ（件／割合）

5分未満	2,716	(29.3%)	1～2時間	714	(7.7%)
5～15分	2,640	(28.5%)	2～5時間	340	(3.7%)
15～30分	1,555	(16.8%)	5～10時間	26	(0.3%)
30分～1時間	1,264	(13.7%)			

表8 相手人数（件／割合）

1人	2人	3人	4人	5人	6人	7人	8人	9人	10人以上
5,257	1,713	932	392	254	129	81	58	40	388
(56.9%)	(18.5%)	(10.1%)	(4.2%)	(2.7%)	(1.4%)	(0.9%)	(0.6%)	(0.4%)	(4.2%)

表9 モード（件／割合）

対面	8,368	(90.3%)
電話・ネット	904	(9.7%)

表10 相手属性（件／割合）

日本人同士	9,188	(99.1%)
外国人含む	84	(0.9%)

表11 言語（件／割合）

日本語	9,200	(99.2%)
外国語（含む）	72	(0.8%)

■**会話状況の属性** 表12～14に会話状況の属性3項目の結果を示す。会話の生じる《時間帯》は、午後の時間が若干多く夜が若干少ないものの、各時間帯に含まれる時間数を考慮すると、どの時間帯もほぼ同程度の会話が生じていると言える（表12）。《場所》については、自宅が35%、職場・学校が30%と、私的あるいは公的に主たる空間で会話が多く生じていることが分かる（表13）。また銀行や商店、レストランなどの公共商業施設での会話も18%と少なからず見られる。《活動》については、仕事・学業が25%、家事・雑事が17%と、家事・雑事を含む仕事・学業の過程で会話を多く交わしていることが分かる（表14）。また食事や休息中の会話も15%前後と少なからず見られる。

² 表6～8は未回答を除く値を示す。

表 12 時間帯 (件 / 割合)

午前	3,019	(32.7%)
午後	3,618	(39.2%)
夜	2,599	(28.1%)

表 13 場所 (件 / 割合)

自宅	3,237	(35.0%)	交通機関	472	(5.1%)
職場・学校	2,802	(30.3%)	それ以外の屋内	411	(4.4%)
公共商業施設	1,700	(18.4%)	それ以外の屋外	631	(6.8%)

表 14 活動 (件 / 割合)

食事	1,566	(17.0%)	仕事・学業	2,331	(25.3%)	付き合い	251	(2.7%)
家事・雑事	1,588	(17.2%)	業務外・課外活動	118	(1.3%)	移動	1,069	(11.6%)
身の回りの用事	501	(5.4%)	社会参加	97	(1.1%)	休息	1,187	(12.9%)
療養	85	(0.9%)	レジャー活動	429	(4.7%)			

以上、項目別の分析から日常会話の傾向をまとめると、次のようになる。

- ① 雑談や用談・相談が多く、会議・会合や授業・レッスン・講演は少ない。
- ② 30 分以下の短い会話が多く、1 時間を超える長い会話は少ない。
- ③ 相手人数が 1 人か 2 人の少人数会話 (2 ～ 3 人会話) が多い。
- ④ 電話などによる遠隔会話は 10% 程度に留まり大半が対面会話である。
- ⑤ 日本人同士の日本語 (のみ) による会話が大半を占める。
- ⑥ 自宅や職場・学校など、私的あるいは公的に主たる空間での会話が多い。
- ⑦ 家事・雑事を含む仕事や学業、食事、休息中の会話が多い。

3.2 調査項目間の関係

本節では、全調査項目のうち、協力者の属性として《性別》《年代》《職業》、会話の属性として《形式》《長さ》《相手人数》、会話状況の属性として《時間帯》《場所》《活動》を取り上げ、項目間の関係を通して会話行動の傾向を見ていく。

3.2.1 項目の値の併合

項目によっては選択肢が多く煩雑となるため、次の方法で項目の値を一部併合した。

まず 9 項目の対応関係を多重対応分析によって分析し、各カテゴリーに与えられた重み係数 (3 次元解) をもとにしたクラスター分析 (ユークリッド距離・Ward 法) を行った。分析には統計解析言語 R の MASS パッケージ中の mca 関数と hclust 関数を用いた。クラスター分析の結果から大きく五つのクラスターに分類できることが分かった (図 2)。

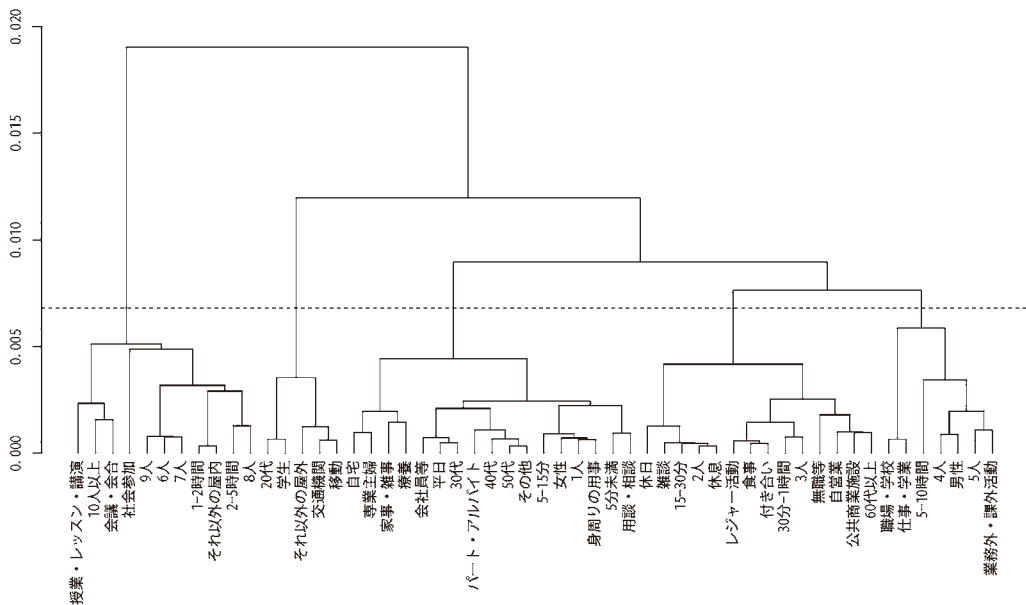


図2 調査項目のクラスター分析の結果

この結果にもとづき、同じクラスターに属する値のうち類似した値を併合することとした。ただし単独で頻度の高い値は併合しない方針とした。例えば《活動》では、「休息」「食事」「付き合い」「レジャー活動」が同じクラスターに属するが、このうちいわゆる積極的レジャーに分類される「付き合い」と「レジャー活動」のみを併合し、単独で高頻度の「食事」や消極的レジャーの「休息」は併合しなかった。また、《職業》の「その他」はいずれも有職者であったため「パート・アルバイト」と併合した。この方針に従い次の通り各項目の値を併合した。なお下線は、次節以降の分析で図中に記載する場合の略称である。冒頭の一文字で略すこともある。

職業：会社員等，自営業，パート等（パート・アルバイト＋その他），学生，専業主婦，無職等
形式：雑談，用談・相談，会議・授業等（会議・会合＋授業・レッスン・講演）

長さ：5分未満 [～5 m]，5～15分 [～15 m]，15～30分 [～30 m]，30分～1時間 [～1 h]，
 1～5時間 [～5 h]，5～10時間 [～10 h]

相手人数：1人，2人，3人，4～5人 [～5人]，6人以上 [6人～]

場所：自宅，職場・学校，公共商業施設，それ以外の屋内，屋外・交通機関（それ以外の屋外＋交通機関）

活動：食事，家事・雑事等（家事・雑事＋身の回りの用事＋療養），移動，休息，社会参加，仕事・学業等（仕事・学業＋業務外・課外活動），レジャー活動等（レジャー活動＋付き合い）

3.2.2 協力者の属性ごとの傾向

本節では、性別・年代別・職業ごとに会話・会話状況の属性6項目の出現傾向を見ていく（図

3～5)。

性別ごとに見た場合(図3)、男性よりも女性の方が、形式については雑談が、長さについては15分未満の短めの会話が、相手人数については2人以下の少なめの会話が若干多く見られる。ただし両者の差は大きくはない。場所と活動については、女性は自宅での家事・雑事等に従事しながらの会話が、男性は職場・学校での仕事・学業に従事しながらの会話が、これは性別による職業の偏り(表5)に起因するものと考えられる。この点については職業別の出現率で改めて触れる。

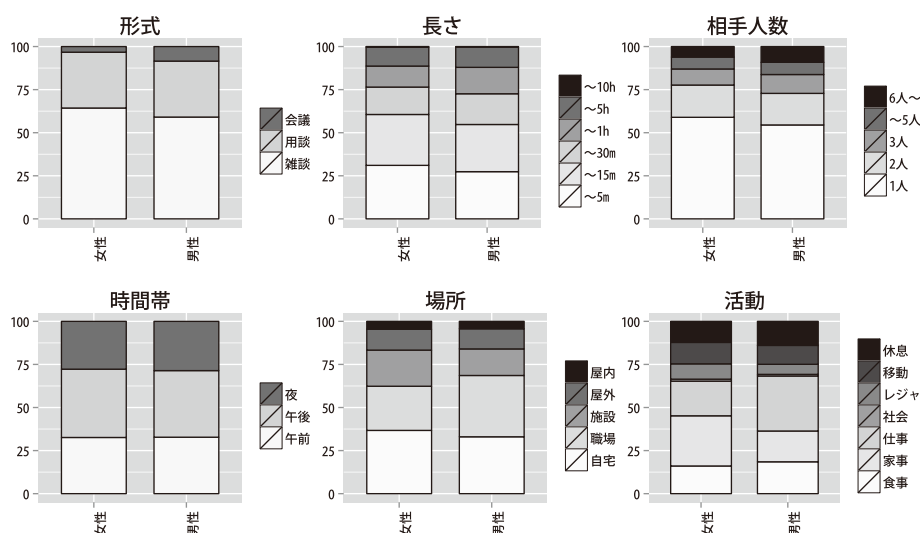


図3 会話の属性・会話状況の属性：性別の出現率

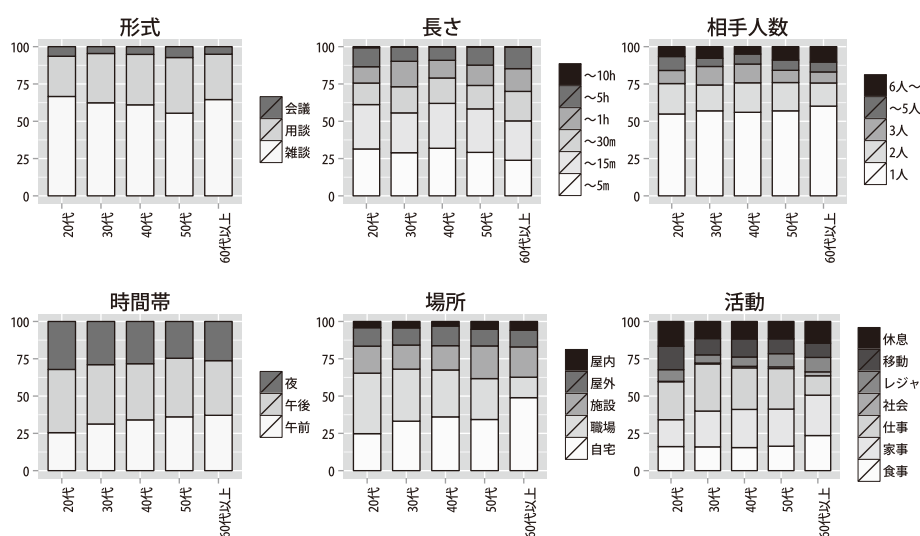


図4 会話の属性・会話状況の属性：年代別の出現率

年代ごとに見た場合も同様に、職業の偏りに起因すると見られる傾向がうかがえる（図4）。例えば、有職率の高い30～50代や学生の多い20代は、職場・学校での仕事・学業中の会話が多く見られるのに対し、定年退職者や専業主婦の多い60代は、自宅での食事や家事・雑事中の会話がが多い。

このように、性別・年代別に見られる会話・会話状況属性の出現傾向は、職業によるところが大きいと考えられる。そこで以下では、職業ごとの傾向について詳細に見ていく（図5）。

形式については、いずれの職業も雑談が全体の半数以上（55.5～72.3%）を占めているのに対し、会議・授業等は2.5～8.3%と出現率は高くない。用談・相談は25.2～36.2%であり、どの職業でも一定数生じていることが分かる。これは前節で見た全体の傾向（表6）と一致する。細かく見ると、会社員等や自営業などの有職者は会議・授業等や用談・相談が多いのに対し、専業主婦は雑談が多い傾向にある。長さについては、自営業を除く全ての職業で15分未満のごく短い会話が全体の半数以上を占めており、1時間以上の長い会話は10～15%程度に留まっていることが分かる³。相手人数は、いずれの職種も1人の場合が全体の半数以上（52.3～60.7%）、2人以下の場合が75%前後（70.1～82.1%）であり、相手が2人以下の少人数の会話が多くを占めるという点において全体の傾向（表8）と一致する。

場所と活動については、職業による違いがかなり見られる。例えば、場所については、専業主婦や無職・定年退職者は自宅や公共商業施設での会話がが多いのに対し、自営業を除く有職者や学生は職場・学校での会話がが多い⁴。活動についても、有職者は仕事中の会話が、専業主婦は家事・雑事中の会話がが多いのに対し、学生は学業に加えて休息や移動中の会話が、無職・定年退職者は食事や家事・雑事に次いで休息やレジャー活動中の会話が多く見られる。

以上、協力者の属性ごとに会話・会話状況属性の傾向をまとめると、次のようになる。

- ① 性別・年代別に見られる会話・会話状況属性の出現傾向は、その性別・年代に属する人の職業の偏りによる可能性が高い。
- ② 会話の形式・長さ・相手人数については、いずれの職業も全体とほぼ同じ傾向を示す。ただし、会話の形式については、有職者は会議・会合や用談・相談が多いのに対し、専業主婦は雑談が多いなどの違いも見られる。
- ③ 場所と活動については、例えば専業主婦は自宅や公共商業施設での家事・雑事中の会話がが多いのに対し、有職者は職場での仕事中の会話が、学生は学校での学業中の会話に加えて休息時や移動中の会話がが多いなど、職業による違いが顕著に見られる。

³ 自営業の場合、短時間の接客などが多いことが予想される。例えば途切れなく接客する場合などで調査の記録が間に合わない場合には、全体をまとめて接客とし合計の時間と人数を報告しても良いとしたため、1時間以上の長めの回答が多くなっていると考えられる。

⁴ 自営業の場合、他の有職者に比べて自宅での会話がが多いが、これは本質的に、職場と自宅が切り分けづらいことに起因すると考えられる。

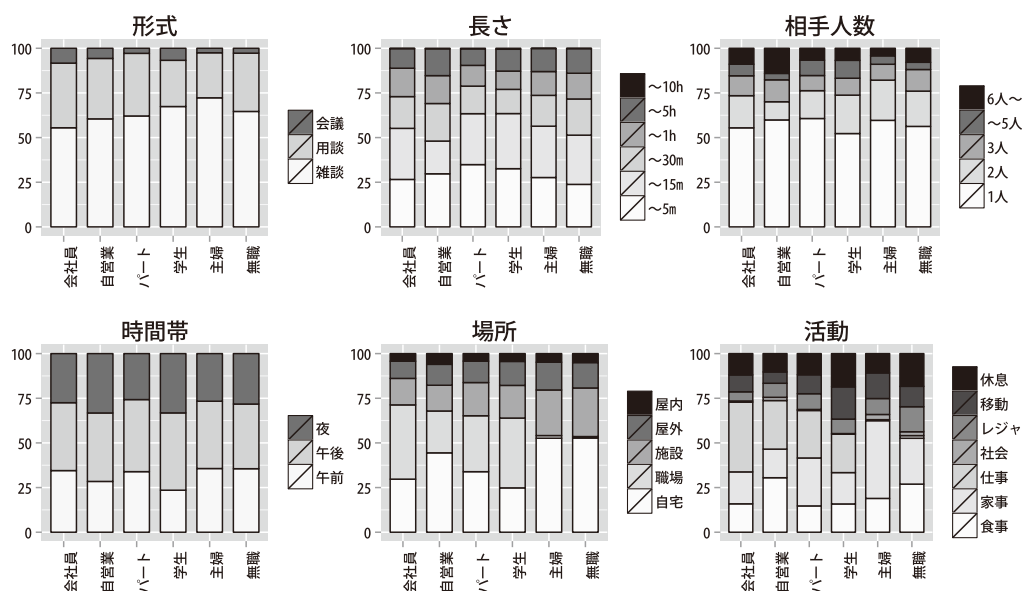


図5 会話の属性・会話状況の属性：職業別の出現率

3.2.3 場所と活動

前節で見たように、場所と活動は職業の違いが反映されやすいため、両者は職業を媒介して対応関係が見られる可能性が高い。そこで本節では、場所と活動に焦点を当て、会話の形式も加えてその対応関係を見ていく。図6(次頁)は、活動の種類ごとに会話の形式と場所の出現傾向を示したものである(「社会参加」は頻度が低いため割愛する)。棒の幅と高さは、形式と場所のそれぞれのカテゴリの出現率を、面積は「形式×場所」のカテゴリの出現率を表している。例えば図中の左上の食事の結果を見ると、雑談が全体の90%(棒の幅)を占めており、そのうちの約60%(棒の高さ)が自宅における食事時の会話であることが分かる。これは、自宅での雑談が食事会話全体の50%以上(0.9×0.6, 棒の面積)を占めていることを意味する。

まず活動と会話の形式との関係に着目すると、食事、移動、休憩では80%以上が雑談であるのに対し、仕事・学業等では56%を用談・相談が、18%を会議・会合が占めており、雑談は26%と少ない。家事・雑事やレジャー活動・付き合いは、雑談が60%強、用談・相談が30~40%であり、食事などと比べて用談・相談が少なからず見られる。

活動と場所との関係を見ると、食事と休息は自宅で、仕事・学業等は職場・学校で、レジャー活動・付き合いは公共商業施設で、移動は屋外で頻出する傾向にある。これに形式の違いを加えて詳細に見てみると、食事は、自宅では雑談が圧倒的に多いが、公共商業施設では雑談に加えて用談・相談も多く見られる。職場での食事では用談・相談よりも雑談をすることの方が多いようである。家事・雑事は、自宅では家族を相手に雑談や用談をしながら行われることが多いのに対し、店舗などの公共商業施設では店員などを相手とする用談が多い。仕事・学業等は圧倒的に職

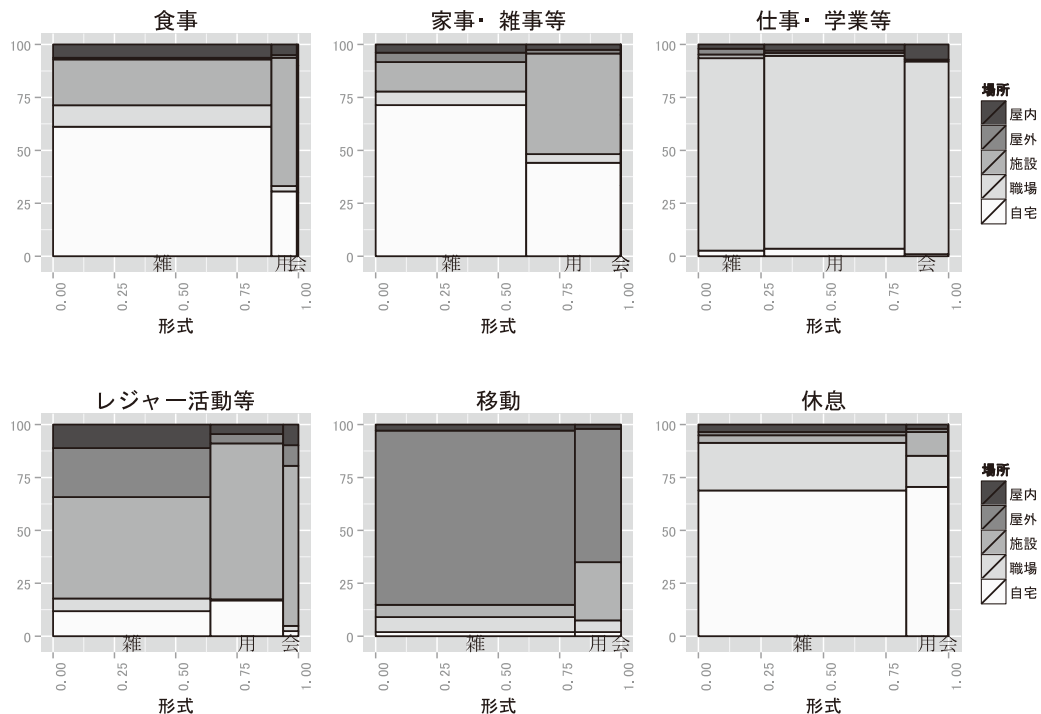


図6 活動ごとに見る形式と場所の出現率

場・学校が多いが、自宅での仕事に会話をすることもある。詳しく見てみると、自宅での仕事・学業中の会話の60%が電話やインターネットなどの遠隔会話であった。全体では遠隔会話は10%に留まることを考えると、ここでの60%という値はかなり高いと言える。レジャー活動・付き合いは公共商業施設が多いが、その傾向は用談・相談の場合に特に顕著に見られる。移動は屋外・交通機関（特に後者）が、また休息は自宅が圧倒的に多い。また、自宅での家事・雑事中の雑談が専業主婦に、職場・学校での仕事・学業中の用談・雑談が会社員等に多いなど、場所・活動と職業との関係もうかがえる。

以上をまとめると、次のようになる。

- ① 食事、移動、休息中は雑談が、家事・雑事やレジャー活動等では雑談に次いで用談・相談が、仕事・学業では用談・相談に次いで会議等が多く見られるなど、活動と会話の形式には対応関係が見られる。
- ② 食事や休息は自宅で、仕事などは職場等で、レジャーなどは公共商業施設で頻出するなど、活動と場所には対応関係が見られる。
- ③ 自宅での家事中の雑談が専業主婦に多いなど、場所・活動と職業との関係も見られる。

3.2.4 会話の人数と長さ

図5で見たように、どの職種においても少人数、短時間の雑談や用談・相談が大半を占める。

しかし、少ないながらも 4 人以上の比較的人数の多い会話や 1 時間以上の長い会話も確実に存在する。こうした会話の人数や長さの点において頻出する事例あるいは稀な事例がどのような状況で生じているかを明らかにするために、会話の形式ごとに会話の人数と長さの関係を見る。

図 7 に、会話の形式ごとに相手人数と長さの出現傾向を示す。図から、用談・相談では 70% 以上が相手人数 1 人であり、その 80% が 15 分未満と、少人数・短時間の会話が頻出しているのに対し、会議・授業等では逆に多人数・長時間の会話が多く見られることが分かる。雑談はその中間の傾向を示している。またいずれの形式においても、相手の人数が増えるほど長い会話が増加する傾向が見られる。

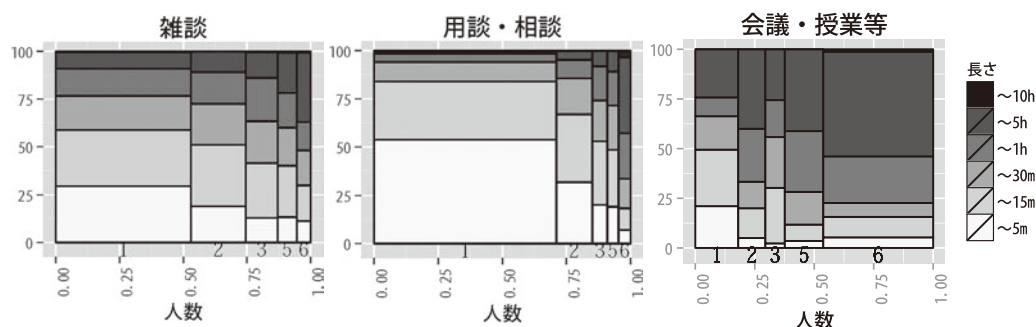


図 7 形式ごとに見る人数と長さの出現率

会話の長さや人数との関係を活動ごとに見ると（次頁図 8）、形式ごとの分析（図 7）で観察された、人数が増えるほど長い会話が増加する傾向が、いずれの活動にも見られることが分かる。

人数に着目して全体の傾向を概観すると、家事・雑事等や移動、休息中は相手が 1 人であることが多いのに対し、食事や仕事・学業等、レジャー活動・付き合いの際は 2 人あるいは 3 人以上の場合が相対的に多い。長さにも着目すると、家事・雑事等は相手が 1 人の場合、5 分未満のごく短い会話が頻出しているのに対し、同じく相手人数の少ない休息は相対的に長めの会話が長く、1 時間以上の会話も 25% 見られる。移動はその中間の傾向を示す。家事・雑事中は、店員や家族と用談を中心とする軽い会話を交わすことはあっても、大人数と長時間話し込んでその活動を止めることはないということであろう。これは、図 7 で見た用談・相談が少人数・短時間の傾向を示すという結果と整合的である。次に、比較的人数の多い食事、仕事・学業等、レジャー活動・付き合いを比べてみると、食事は長めの会話が長いのにに対し、仕事・学業等では短めの会話が長い。レジャー活動等はその中間の傾向を示す。仕事・学業等に短時間の会話が長いのは家事・雑事等と同じ理由であると考えられるが、前者は会議などの大人数・長時間の会話も見られる点に違いがある。また、食事とレジャー活動・付き合いは、ある程度まとまった時間の会話をするこの多い活動と言える。特にその傾向は食事に強く見られる。

そこで、食事に焦点を当て、場所として自宅、職場・学校、公共商業施設を取り上げ、相手人数と長さの関係を見てみよう（次頁図 9）。自宅では、相手が 1 人の少人数会話が半数を占めているが、これは特に午前（朝）の慌ただしい時間帯にこの傾向が強く見られ、夜は人数の比較的

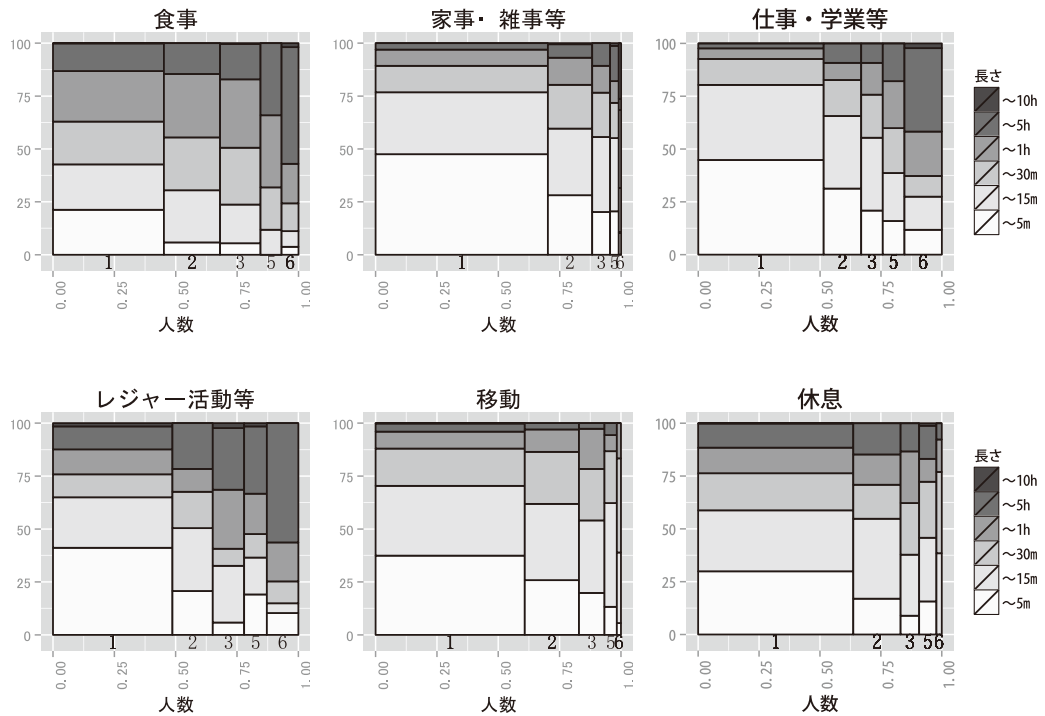


図8 活動ごとに見る人数と長さの出現率

多い長い会話が見られる。職場・学校では、相手が3人以上の比較的大人数の会話が半数を占めており、また30分以上の会話も46%を占めているなど、複数の同僚や友人と職場・学校で食事しながらまとまった時間をかけて会話する機会が多いことがうかがえる。レストランなどの公共商業施設では、1時間を超えるかなり長い会話が多いのが特徴である。相手1人・5分以下の会話も目立つが、これは注文時の店員とのやりとりである。

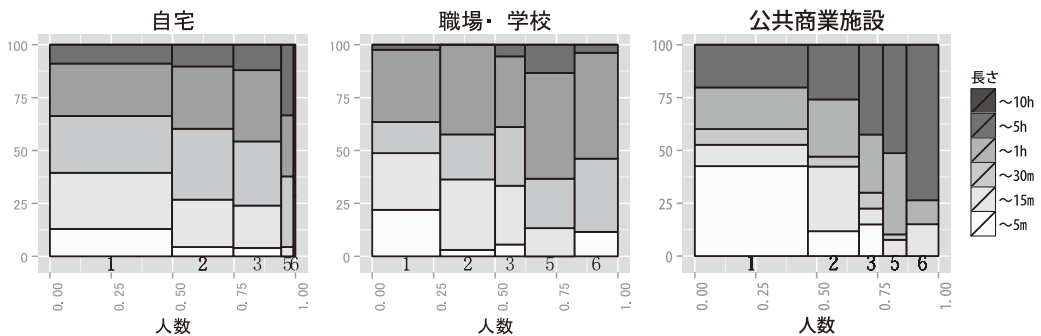


図9 場所ごとに見る人数と長さの出現率（食事の場合）

以上をまとめると、次のようになる。

- ① 用談・相談は少人数・短時間の会話が、会議・授業等は多人数・長時間の会話が多い。雑

談はその中間の傾向を示す。

- ② 食事中は、ある程度の人数とまとまった時間の会話をする事が多いのに対し、家事・雑事や仕事・学業中の会話は、当該活動に支障をきたさない短時間会話が早いといったように、会話の人数・長さや活動の種類との対応関係が見られる。
- ③ 相手の人数が増えるほど長い会話が増加するという一般的な傾向が見られる。

3.2.5 会話相手の属性

これまでは、会話相手の属性に関わらずその合計人数だけに着目してきた。本節では、どのような関係の人と会話をする傾向にあるかを見ていく。全会話の全ての話者数を関係性ごとにまとめ上げ、形式あるいは活動の種類ごとにその出現率を求めた。図 10 に結果を示す。

形式ごとの傾向を見てみると（図 10 左）、雑談の場合、家族や友人知人、仕事学業関係といったように、私的あるいは公的に普段からよく接している人が多いのに対し、用談・相談は仕事学業関係者や公共商業関係が、会議・会合や授業・レッスン・講演は先生生徒や仕事学業関係が大半を占めるなど、公的な関係の人や公共商業関係の人が多い。これらの形式では、雑談と比べ、顔見知り・見知らぬ人と会話することも多い。

次に活動ごとの結果を見てみると（図 10 右）、家族との会話が多いのは、食事や家事・雑事等、休息中であり、また移動時も少なからず見られる。友人・知人との会話も家族の場合と類似した傾向を示すが、家事・雑事では少なく、むしろ社会参加（地域活動やボランティアなど）やレジャー活動・付き合いでの会話が多くなる。仕事・学業関係者の場合、仕事・学業中の会話が大半を占めるが、食事や出勤・退勤時の移動、また休息（休憩時間）中の会話も少なからず見られる。顔見知り・見知らぬ人との会話は、地域活動などの社会参加やレジャー活動中などに見られるようである。

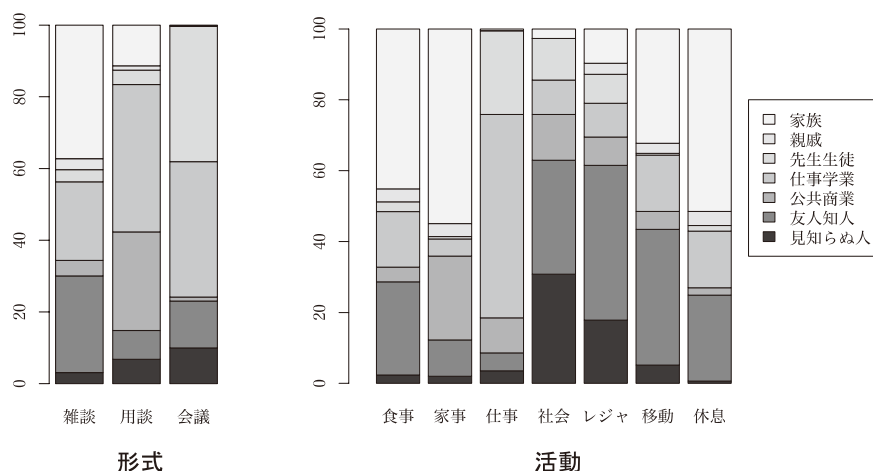


図 10 形式・活動別に見る相手属性クラスの出現率

以上をまとめると、次のようになる。

- ① 雑談は私的あるいは公的に普段接している人が多いのに対し、用談・相談や会議・会合は公的な関係の人や公共商業関係の人が多い。
- ② 家族と友人・知人は食事や休息中など共通した活動中に会話することが多い。
- ③ 仕事学業関係者は、本業中のほか、②と同様、食事中・休息中の会話も多い。
- ④ 顔見知り・見知らぬ人との会話は、社会参加やレジャー活動中に見られる。形式としては用談・相談や会議・会合が多い。

4. コーパスの設計に向けて

4.1 収録法

設計について考える前に、現在検討している日常会話の収録法について言及する。

まず、過去に均衡性を考慮して構築された話し言葉コーパスについて概観しておこう。話し言葉コーパスのうち、均衡性を考慮して構築されたものはほとんど存在しないのが現状である。これはまさに、話し言葉のレジスターをいかにとらえるかが問題となっていることの現れである。そのような中、British National Corpus (BNC) は、話し言葉について（も）均衡性を考慮して設計された数少ないコーパスである。

BNC は、1990 年代前半にイギリスで構築された、イギリス英語を対象とする 1 億語からなる大規模コーパスである。大半は書き言葉から構成されるが、1 割に当たる 1 千万語が話し言葉に割り当てられている。話し言葉のパートは、次の二つのデータ群から構成される (Crowdy 1995, Burnard and Aston 1998)。

1. 年齢・性別・社会クラス・地域に偏りがないよう選ばれた 124 人のインフォーマントが 7 日間にわたり自身で収録した日常会話を対象とするデータ群
2. 放送や講義など公的かつ重要な（受け手の多い）場での話し言葉を、教育・教養、ビジネス、団体、レジャーの四つの領域に分けて収集したデータ群

後者には、授業での教師と生徒のやりとりやビジネス場面での会合など、本調査で授業・レッスン・講演や会議・会合に分類されるものが含まれている。

BNC が特徴的なのは、日常場面の中で自然に生じた会話を収録しているところである。前節の調査で見てきたような多様な日常場面での会話を収録するには、収録場面を人工的に設定するのではなく、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) を対象としなければならない。上記のうち特に 1 に示す方法は、こうした自然発生的な会話の収録にふさわしいものと言える。我々もこの方針を踏襲し、次の二つの方法でデータを収録することを計画している。

1. **個人密着法**：性別・年代などの点から均衡性を考慮して選別されたインフォーマントに収録

機材等を1ヶ月程度貸し出し、インフォーマント自身に日常会話を収録してもらう方法。調査者は原則として介在しない。

2. **特定場面法**：職場・学校での会議や授業、店舗での店員とのやりとりなど、個人密着法では技術的・倫理的に収録が難しいと思われる場面を特定し、調査者が主体となり収録する方法。調査者は介在するが、日常場面の中で自然に生じる会話を対象とする。

4.2 インフォーマント

個人密着法では、インフォーマントが主体的に関わる会話を収録することになるため、インフォーマントの属性を適切に設定する必要がある。基本的に、会話や会話行動の多様性を考えた場合、生年代と性別の均衡性の考慮は不可欠であろう。今回の行動調査のように、5世代（20代、30代、40代、50代、60代以上）の男女の数をある程度均等に割り振る方法などが考えられる。また今回の調査で、インフォーマントの職業が会話の形式や場所、活動の種類に強く影響することが明らかになったため、職業の点においても偏りがないう配慮する必要がある。このように話者の属性を偏りなく配分することによって、結果として多様な会話を収録することが可能となる。

4.3 会話・会話状況属性の比率

個人密着法のように、自然に生じる会話を収録する方法を採用する限り、収録する時点で会話の形式や人数、長さ、活動、場所などを統制することは難しい。実際、BNCの1のデータ群も、インフォーマントの属性（年齢・性別・社会クラス・地域）の点では均衡性を確保しているが、会話や会話状況の属性については均衡性を考慮しておらず、偏っている可能性もある。またBNCの2のデータ群は、こうした会話や会話状況の偏りを解消するために収録されたものではなく、あくまで放送や講義など公的かつ重要な（受け手の多い）場での話し言葉を補強するために設計されたものであり、またその構成比率も十分練られたものではないとしている。

我々は、今回の会話行動調査の結果を参考に各項目の比率を算出し、それを一つの目安に、収録された各種会話からコーパスとして整備・公開する対象を選別することを考えている。個人密着法では、雑談と用談・相談が、特定場面法では会議・授業等が収録の中心となるだろう。そのため、まずは会話の形式でコーパス全体をいかに配分するかを決める。前節の結果を見ると、雑談と用談・相談は、いずれの協力者属性で見ても、前者が6割前後、後者が3割前後の出現率となっている。また会議・授業等の出現率は職業によって偏るが、最も多い有職者や学生に限定しても、両者合わせて1割に満たない。こうした比率がコーパス設計の一つの指標となる。

ただし、これは会話の出現回数で見た場合の値である。3.2.4節で見たように、会議・授業等は長く、用談・相談は短い傾向にあることから、長さで見ると比率は異なってくる。そこで、調査項目の会話の長さからその平均を用いて各形式の比率を算出したところ⁵、雑談が7割弱、用談・

⁵ 例えば「30分～1時間」の場合はその平均をとって「45分」とした。

相談が2割弱、会議・授業等が1.5割弱であった。

以上、会話行動調査の結果に基づき、会話の件数・長さの観点から各形式の比率を配分すると、表15のようになる。

表15 会話コーパスの設計案：形式の比率の目安

	形式		
	雑談	用談・相談	会議・授業等
会話の件数	6割前後	3割前後	1割
会話の長さ	6.5～7割	2割	1～1.5割

次に問題となるのが個々の形式をいかに分けるかということである。収録時の統制が難しい以上、全ての項目について細かく数値設定をしても、実現性に乏しいものとなるだけである。そこで、主な項目について大雑把にその比率を考察しよう。前節で示した調査結果をまとめると次のようになる。

- ① 自宅や職場・学校など、私的あるいは公的に主たる空間での会話が多い（3.1節）。
- ② 活動や場所には、職業との対応関係が見られる（3.2.2節、3.2.3節）。
- ③ 活動と形式、活動と場所には対応関係が見られる（3.2.3節）。
- ④ 会話の人数や長さには、形式との対応関係が見られる（3.2.4節）。
- ⑤ 会話の人数や長さには、活動との対応関係が見られる（3.2.4節）。

会話の人数や長さについては、会話の形式や活動との対応関係があることから（④・⑤）、形式や活動さえバランスよく配分すれば、おのずと人数や長さもバランスよく分散することが期待される。よって、会話の人数や長さの比率については、必ずしもコーパス設計に盛り込む必要はない。一方、会話の人数や長さに影響を与える活動はどうだろうか。活動と場所是对応関係が見られることから（③）、両者を合わせて考えてみよう。

いずれもカテゴリーが多いため、類似したカテゴリーをまとめる必要がある。場所については、①にあるように、基本的には自宅などの私的な空間、職場や学校などの公的な空間が多いことから、「私的空間」「公的空間」「それ以外」の三つに分けてとらえることができる。活動については、食事と休息は形式や場所、会話の長さや人数の比率が類似しており（図6・図8）、調査項目のクラスター分析でも同じクラスターに分類されていることから（図2）、一つにまとめられる。一方、家事・雑事等と仕事・学業等は、前者が家や公共商業施設、後者が職場・学校に多く見られるなど違いはあるものの（図6）、職業として前者は専業主婦が、後者は会社員等の有職者や学生といったように、職業との対応関係が見られることから（②）、インフォーマントの職業の均衡性さえ考慮すれば結果として家事・雑事等と仕事・学業等はバランスよく分散することが予想される。よって、これらをまとめ上げると、活動は、「食事・休息」「仕事・学業」「その他」の三つに分けることができる。

上記のように整理した上で、三つの形式ごとに、活動と場所の件数と長さの比率を求めると、表16のようになる。

表 16 会話コーパスの設計案：形式内の活動・場所の比率の目安

	活動 場所	形式								
		雑談			用談・相談			会議・授業等		
		私的	公的	それ以外	私的	公的	それ以外	私的	公的	それ以外
会話の 件数	食事・休息	30%	5%	10%	5%		5%	5%		
	仕事・学業	15%	10%	5%	15%	40%	20%	0%	70%	10%
	その他	5%		20%	0%	0%	15%	0%	0%	15%
会話の 長さ	食事・休息	30%	5%	20%	10%		5%	5%		
	仕事・学業	10%	5%	5%	10%	50%	15%	0%	70%	10%
	その他	5%		20%	0%	0%	10%	0%	0%	15%

これらの数値を目安に、収録された各種会話からコーパスとして整備・公開する対象を選別すれば、均衡性を有する日常会話コーパスを構築できる。ただし、最終的な配分は収録の技術的・倫理的な問題なども総合的に考慮して決める必要があるだろう。

5. おわりに

本稿では、大規模な日本語日常会話コーパスの構築を目指し、その基盤整備の一環として実施した、コーパス設計のための会話行動調査について報告した。またこの調査を受け、コーパス設計の方針について議論した。4節においてコーパス設計案として示した各項目の比率は、あくまで今回の調査のみにもとづき算出したものであり、今後、収録の技術的・倫理的な問題なども総合的に考慮して見直す必要がある。このように、まだ検討すべき課題は数多く残されているが、世界的に見ても、こうした調査に基づき日常会話をバランスよく収集することを目指したコーパスは存在しない。実現に向けて着実に準備を進めたい。

参考文献

- Burnard, Lou and Guy Aston (1998) *The BNC handbook*. Edinburgh: Edinburgh University Press. (北村裕 (監訳) (2004) 『The BNC handbook : コーパス言語学への誘い』 東京 : 松柏社).
- Crowdy, Steve (1995) The BNC spoken corpus. In: Geoffrey Leech, Greg Myers, and Jenny Thomas (eds.) *Spoken English on computer: Transcription, mark-up and application*, 224-235. Harlow: Longman.
- 伝康晴・榎本美香 (2014) 『千葉大学3人会話コーパス』 使用説明書 Release1』 <http://research.nii.ac.jp/src/Chiba3Party.html>
- 伝康晴・フライ ジョン (2000) 「CallHome Japanese Corpus」『音声研究』 4(2): 24-30.
- 現代日本語研究会 (編) (1999) 『女性のことば・職場編』 東京 : ひつじ書房.
- 現代日本語研究会 (編) (2002) 『男性のことば・職場編』 東京 : ひつじ書房.
- 畠弘巳 (1983) 「場面とことば」『国語学』 133: 55-68.
- 小磯花絵・土屋智行・渡部涼子・横森大輔・相澤正夫・伝康晴 (2015) 「均衡会話コーパス設計のための一日の会話行動に関する調査—中間報告—」『第7回コーパス日本語学ワークショップ予稿集』 27-34.
- 国立国語研究所 (1971) 『待遇表現の実態—松江24時間調査資料から—』 (国立国語研究所報告 41) 東京 : 秀英出版.
- 国立国語研究所 (1980) 『日本人の知識階層における話しことばの実態 : 「場面について」 分析資料』 東京 : 国立国語研究所 (内部資料).
- 国立国語研究所 (1987) 『談話行動の諸相 : 座談資料の分析』 (国立国語研究所報告 92) 東京 : 三省堂.
- Maekawa, Kikuo, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae

- Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den (2014) Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation* 48(2): 345–371.
- 日本放送協会 (2010) 『2010 年国民生活時間調査報告書』東京：NHK 放送文化研究所.
- 宇佐美まゆみ (2007) 『改訂版：基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ)』2007 年 3 月 31 日改訂版』 <http://www.tufs.ac.jp/ts/personal/usamiken/btsj.htm>

Survey of Conversational Behavior: Towards the Design of a Balanced Corpus of Conversational Japanese

KOISO Hanae^a TSUCHIYA Tomoyuki^b WATANABE Ryoko^c YOKOMORI Daisuke^b
AIZAWA Masao^d DEN Yasuharu^e

^aDepartment of Linguistic Theory and Structure, NINJAL

^bKyushu University

^cAdjunct Researcher, Center for Corpus Development, NINJAL

^dDepartment of Language Change and Variation, NINJAL

^eChiba University / Invited Professor, Department of Corpus Studies, NINJAL

Abstract

In this paper, we report on the activity of a preparatory project to build a large-scale corpus of conversational Japanese (NINJAL collaborative research project, 2014/7/1–2015/8/31). The overall aims of this project are: i) to establish a corpus design for collecting various kinds of everyday conversations in a balanced manner, ii) to develop a methodology of recording naturally occurring conversations, and iii) to create a transcription system suitable for effectively transcribing natural conversations. This report focuses on the first issue of establishing a corpus design. We first describe our survey of everyday conversational behavior, conducted with about 250 Japanese adults last year, in order to reveal how diverse our everyday conversational behavior is, and to build an empirical foundation for corpus design. The questionnaire included when, where, how long, with whom, and in what kind of activity informants were engaged in conversations. We found that ordinary conversations show the following tendencies: i) they mainly consist of chats, business talks, and consultations; ii) in general, the number of participants is small and the duration is short; iii) many conversations are conducted in private places such as homes, as well as in public places such as offices and schools; and iv) some questionnaire items are related to each other. Based on these results, we discuss how to design a balanced corpus of conversational Japanese.

Key words: survey of conversational behavior, corpus design, corpus of conversational Japanese