

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—

小磯 花絵, 土屋 智行, 渡部 涼子 (国立国語研究所), 横森 大輔 (九州大学),
相澤 正夫 (国立国語研究所), 伝 康晴 (千葉大学/国立国語研究所)

Interim Report on the Survey of Conversational Behavior: Towards the Design of Balanced Corpus of Conversational Japanese

Hanae Koiso, Tomoyuki Tsuchiya, Ryoko Watanabe (NINJAL),
Daisuke Yokomori (Kyushu University), Masao Aizawa (NINJAL),
Yasuharu Den (Chiba University / NINJAL)

要旨

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」では、大規模な日本語日常会話コーパスの構築を目指し、均衡性を考慮したコーパスの設計案の策定を進めている。その準備段階として、一日の会話行動の種類と従事時間に関する調査を行っている。調査では、首都圏在住の成人 200~250 人を対象に、任意の平日2日・休日1日(合計3日/人)の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。本稿では、調査の中間結果について報告すると同時に、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、その方針について議論する。

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求められている。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世紀初頭の日本人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・伝承することは、この時代に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」(代表:小磯花絵, 2014年7月~2015年8月)では、21世紀初頭の日本人の多様な会話行動を納めた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会話コーパスの設計の策定を進めている。

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捉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、生産実態と流通実態を反映したサブコーパスを設計した(Maekawa et al. 2014)。一方、日常会話コーパスの対象として予定しているのは、家庭や地域、職場、学校などで我々が直接交わす会話で

あり、いわゆるメディアを通じて受信する第三者による会話は対象としない。そこで本プロジェクトでは、我々が日常的に直接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、現代日本人が日常、どのような種類の会話をどの程度行っているかを調査することとした。ただし、日常会話の生産実態（構成比）をそのままコーパス設計に反映させるのではなく、まずは日常会話にどのようなレジスター的多様性があるかとらえ、多様な日常会話を網羅したコーパスを設計することを目指す。

本稿では、現在進行中の調査の概要と中間結果について報告した上で、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、その方針について議論する。

2. 会話行動調査

2.1 調査項目の設計

会話行動を調査するにあたり、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。本研究では会話行動を大きく、(1) 調査協力者（以下、協力者）の属性（性別や年代、職業など）、(2) 会話の属性（会話の形式や会話の長さなど）、(3) 会話状況の属性（会話の行われた時間帯や場所、活動など）の三つの軸でとらえることとした。この方針に従い表1から表3に示す調査項目を設計した。以下ではこのうち**形式**、**場所**、**活動**について簡単に補足する。

■**形式** 「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。以上を参考に本調査では、会話の目的をもたない「おしゃべり」「あそび」を「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「教え・さしず」は「授業・レッスン・講演」とした。

■**場所** 「公共商業施設」には、公的に物品やサービスなどをやりとりする場として公共施設（市役所や銀行など）と商業施設（遊園地や店舗など）を含めた。「それ以外の屋内」には取引先の会社や知人、親戚の家などが、「それ以外の屋外」には公園や遊歩道などが該当する。「職場・学校」は協力者の職業から特定できるため一つにまとめた。この分類は国立国語研究所(1980)で述べられている「会話場面」を参考に一部変更して設計した。

■**活動** 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）にもとづき設計した。このうち会話行動を伴わない「睡眠」や他活動と共起して現れる「マスメディア接触」は除外した。また「通勤」「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」「学業」は「仕事・学業」にまとめた。ただし、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。

その他の選択肢として、「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」は入浴や着替え、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

表1 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

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
相手属性	関係ごとの人数	数値入力式	
モード	外国人を含む会話	オプション式	(該当する場合に選択)
言語	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)
	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	会話をした場所	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

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該当する。

2.2 調査の方法

言語生活の記録という意味では地理的多様性を無視することはできない。しかし、時間的・予算的な制約もあることから、第一段階として首都圏にしぼってコーパスを設計・構築することとした。このようにコーパスの対象を首都圏に限定したため、調査対象も首都圏在住者とした。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、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。調査の概要を以下に示す。

■**目的** 日常会話の多様性を明らかにし、それに立脚して多様な日常会話を網羅したコーパスを設計するために、首都圏在住者を対象に1日の会話行動の種類や時間などを調査する。

■**期間** 2014年11月1日~2015年2月末(予定)

- 調査日・時間** 任意の平日2日・休日1日(計3日/1人)の起床から就寝まで
- 対象** 首都圏(東京・神奈川・千葉・埼玉)在住の20歳以上の日本語母語話者200~250人(20代・30代・40代・50代・60代以上×男・女×20~25人)。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。
- 調査項目** 協力者の属性(表1の5項目), 会話・会話状況に関する調査項目(表2の6項目, 表3の3項目), および参考のために会話の概要(自由記述)。
- 手続き** (1) 協力者に調査の手引きと調査票(1日1冊, 計3冊)など資料一式を事前に郵送した。(2) 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)を選択して調査を実施した。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。(3) 調査日当日, 協力者は調査票を携帯し, 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・表3)に回答した。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。(4) 調査終了後, 協力者は調査票と協力者の属性(表1の5項目)を記したシートを調査者に返送した。
- 謝礼** 3日間の調査に対し6000円。

3. 調査結果の分析

3.1 方法

2015年1月28日現在の有効回答219名分(計657日, 8296会話)を対象に分析を行う。219名の内訳を表4・5に示す。なお, 1日の平均会話数は12.6(平日:13.2, 休日:11.4), 1日の平均会話時間は6.1時間(平日:6.0時間, 休日:6.5時間)である*1。また世帯員数は, 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名である。

表4 調査対象：性別・年代の内訳

	20代	30代	40代	50代	60代以上
女性	23	24	25	24	25
男性	15	20	17	21	25

表5 調査対象：職業の内訳

会社員・役員・公務員・専門職	87	自営業	10
パート・アルバイト	32	学生	28
無職・定年退職者	18	その他	12
専業主婦	32		

本研究では, 表1~表3のうち, **世帯員数**, **居住地**, **相手属性**, **モード**, **言語**を除く9項目を分析に用いた。**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。

分析に際し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, 各値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。クラスター分析の結果から大きく5つのクラスターに分類できることが分かった(図1)。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併合した。

*1 会話時間は, **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と予想される。

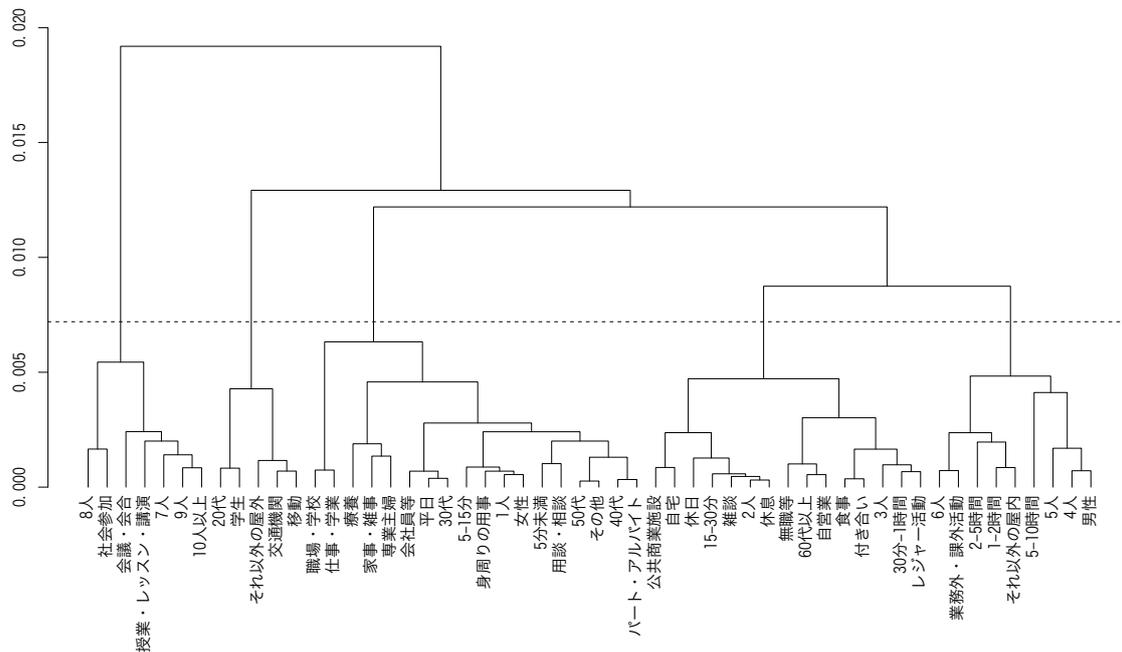


図1 調査項目のクラスター分析の結果

- 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, 5~15分, 15分~1時間 (15~30分 + 30分~1時間), 1時間以上 (1~2時間 + 2~5時間 + 5~10時間 + 10時間以上)
相手人数 1人, 2人, 3人, 4~6人, 7人以上
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), 仕事・学業, 社会参加等 (社会参加 + 業務外・課外活動), レジャー活動等 (レジャー活動 + 付き合い), 移動, 休息

併合した項目を含め, 協力者の属性3項目 (性別, 年代, 職業), 会話の属性3項目 (形式, 長さ, 相手人数), 会話状況の属性3項目 (時間帯, 場所, 活動) を分析に用いた。

3.2 結果

会話の属性3項目の出現傾向を職業別に見てみよう (図2上段)。**形式**については, いずれの職業も雑談が全体の5割以上 (55.6~71.1%) を占めているのに対し, 会議・授業等 (会合やレッスン, 講演など含む) は, 有職者や学生が若干多い傾向を示すものの, 2.7~7.6% と出現率は高くない。用談・相談は24.4~36.9%であり, どの職業でも一定数生じていることが分かる。**長さ**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*2。**相手人数**は, 1人の場合が全体の5割以上 (54.3~65.4%), 3人以下の場合が全体の83.1~91.1%を占めており, 日常会話の大半が少人数の会話であることが分かる。

*2 自営業の場合, 短時間の接客などが多いことも予想されるが, 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あわない場合には, 全体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, 正確な値になっていない可能性がある。

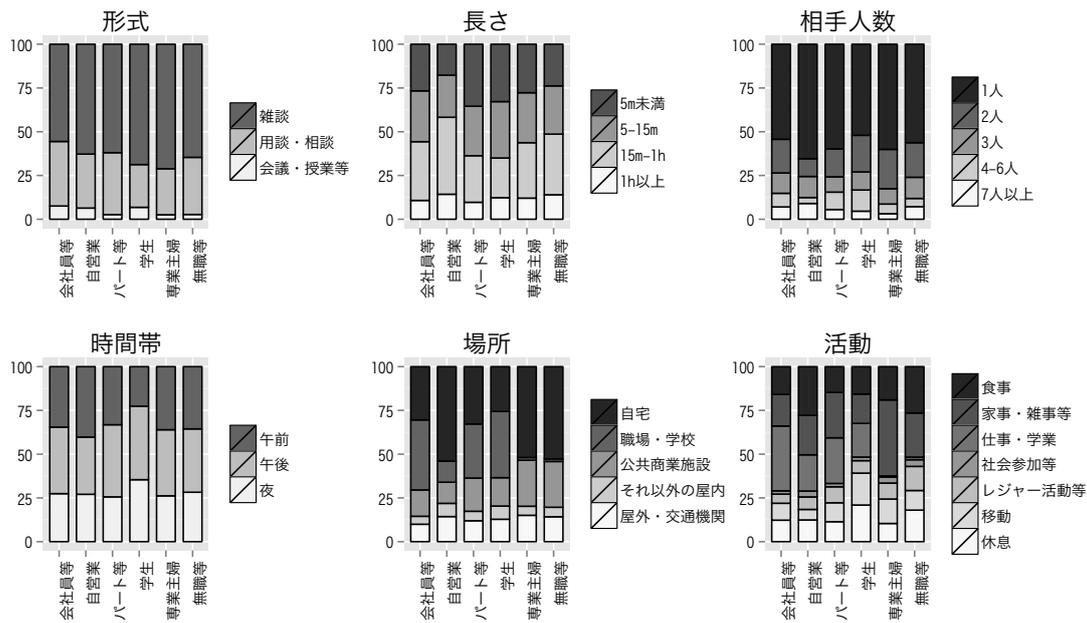


図2 会話の属性・会話状況の属性：職業別に見た出現率 (%)

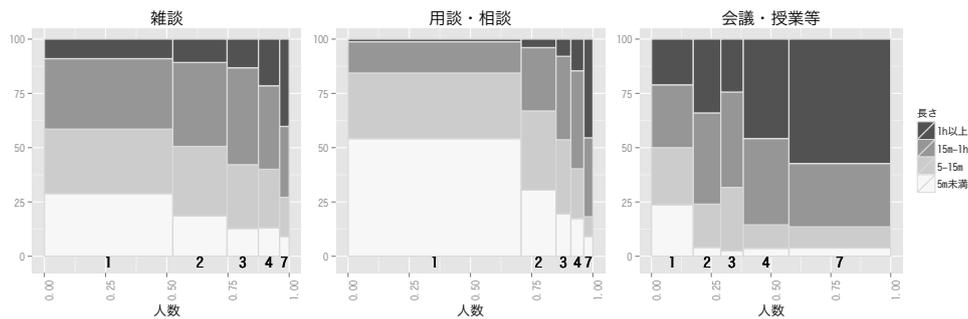


図3 会話の属性3項目間の関係 (横軸ラベル 1: 1人, 2: 2人, 3: 3人, 4: 4-6人, 7: 7人以上)

このように会話の3属性については、総じて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傾向にあると言える。結果は省略するが、性別ごと、年代ごとに見ても同じ傾向がうかがえる。しかし、少ないながらも会議・授業等や4名以上の比較的人数の多い会話、1時間以上の長い会話も確実に存在する。こうした頻度の少ないケースがどのような状況で生じているかを見るために、会話の3属性間の関係を見てみよう。

図3に相手人数と長さの出現傾向を形式ごとに示す。スロットの幅と高さは相手人数と長さのそれぞれの出現率を、面積は相手人数×長さのカテゴリーの出現率を表している。図から、用談・相談では相対的に少人数・短時間の会話が頻出するのに対し、会議・授業等では多人数・長時間の会話が多く見られる。雑談はその中間の傾向を示す。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。例えば、高頻度の典型的な事例として相手1人の5分未満の用談・相談を協力者の3属性ごとに見てみると、いずれの場合も全体の10%前後を占めており、相手1人の5分未満の用談・相談がどの属性の人に

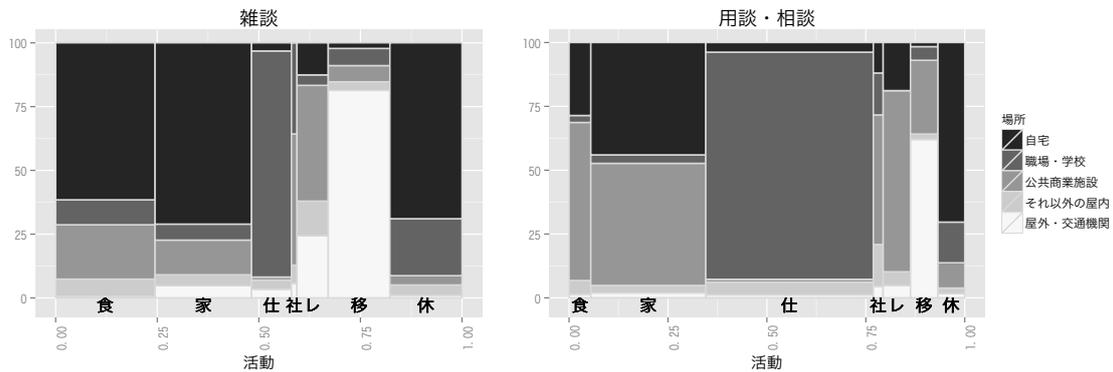


図4 場所と活動の出現傾向 (横軸ラベル 食:食事, 家:家事・雑事等, 仕:仕事・学業, 社:社会参加等, レ:レジャー活動等, 移:移動, 休:休息)

とっても典型的な事例であることがうかがえる。

次に会話状況の属性3項目の出現傾向を見てみると(図2下段), **場所と活動**については, 例えば自宅での会話や家事・雑事中の会話が専業主婦に, 職場・学校での仕事・学業中の会話が会社員等に多いなど, 職業による違いがかなり見られる。また**場所と活動**の関係を会話の**形式**ごとにみると(図4, スペースの都合で出現率の低い会議・会合等は省く), 雑談は, 自宅での食事や家事・雑事, 休息中, 職場・学校での仕事・学業中, 屋外・交通機関(特に交通機関)での移動中に, 用談・相談は, 職場・学校での仕事・学業中や自宅・公共商業施設での家事・雑事中に多く見られるなど, **場所と活動**の対応関係が見られる。また, 図は省略するが, 例えば自宅での家事・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・相談が会社員等に多いなど, **場所・活動**と協力者の属性との関係もうかがえる。

4. コーパス設計に向けて

British National Corpus (BNC) の話し言葉のパートは, 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日間にわたって自身で収録した日常会話を対象とするデータ群と, 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から構成される(Crowdy 1995, Burnard and Aston 1998)。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(分析では会議・授業等)に分類されるものが含まれている。

我々もBNCと同様, 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を一つの柱としつつ, 会議や授業など, この方法では収録が難しいであろう場面を個別に収録する方法も並行して行うことを計画している。前者は雑談と用談・相談が, 後者は会議・授業等が中心となる。よって, 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ことになるだろう。調査はまだ完了していないが, 今回の中間結果を見ると, 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・授業等の会話の出現率は**職業**によって偏るが, 有職者や学生に限定すると1割弱である(図2)。こうした割合がコーパス設計の一つの指標となりうる。ただし, これは会話の回数で見た場合であり, 会議・授業等は長く用談・相談は短い傾向にあることから(図3), 長さで見ると比率

は異なる。回数と長さの両面から調査して比率を決める必要がある。

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。雑談や用談・相談と比べてより単純な構成の会議・授業等から見てみよう。図4ではスペースの都合で省略したが、会議・授業等では、当然のことながら有職者や学生による職場・学校での会議・授業がその大半を占めている。しかしそれに加えて、ボランティアや地域活動など社会参加中の会合や趣味・教養の教室でのレッスンなども少なからず見られ、また協力者の属性も異なることから、会話の多様性を確保する上で重要なレジスターであることが分かる。

このように**活動**と**場所**は、協力者の属性と連動する傾向にあり、結果として会話のレジスターに強く影響することが予想される。この傾向が雑談や用談・相談にも見られることは前節で言及した通りである。もう少し詳しい分析が必要だが、**形式**内の配分は、まずは**活動**か**場所**のいずれかあるいは両方を参考に決めることになる。

また、雑談や用談・相談の大半はインフォーマント（協力者）自身が収録する方法で集めることを検討しているため、多様なレジスターの会話を確保するには、インフォーマントの属性をいかに設定するかが重要となる。**性別**と**世代**のバランスは必須として、上で議論したように、**職業**は特に**場所**や**活動**に大きく影響することから、**職業**も含めてインフォーマントの配分を設定することが求められる。しかし現時点では、インフォーマントは50人程度を見込んでおり、性別×5世代ごとに4~5人となると、職業をどのように組込むかが課題となる。

今回は触れなかったが、電話・ネットでの会話も全体の1割ほどを占めており、コーパスの設計に組み込む必要がある。また倫理的・技術的な観点からの収録の可否も加味しなければならない。小磯ほか(2015)では、各調査項目間の関連をアソシエーション分析によって抽出し、例えば1時間以上の長い会話には授業や会議・会合以外にも食事中や女子学生の雑談が見られるなど、複数項目間にわたって見られる傾向も明らかにしている。今後、さまざまな観点からの分析や検討を重ね、コーパス設計の策定を進めたい。

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』 松柏社,).
- Crowdy, S. (1995). "The BNC spoken corpus." G. Leech, G. Myers, and J. Thomas (Eds.), *Spoken English on computer: Transcription, mark-up and application*. Harlow: Longman. pp. 224–235.
- 畠弘巳 (1983). 「場面とことば」 『国語学』, 133, pp. 55–68.
- 小磯花絵・伝康晴・土屋智行・渡部涼子・横森大輔・相澤正夫 (2015). 「一日の会話行動に関する調査とその準備的分析—均衡会話コーパス設計に向けて—」 『言語処理学会第21回年次大会発表論文集』. 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: 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』: 国立国語研究所日本語教育センター.
- 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: 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, pp. 345–371.
- 日本放送協会 (2010). 『2010年国民生活時間調査報告書』: NHK放送文化研究所.