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—

小磯 花絵 * 伝 康晴 †,* 土屋 智行 * 渡部 涼子 * 横森 大輔 ‡ 相澤 正夫 *
* 国立国語研究所 † 千葉大学文学部 ‡ 九州大学大学院言語文化研究院

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は重要な課題である。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，ことば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社会生活を知るための貴重な記録となる。民族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と言えよう。

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均衡性を考慮した会話コーパスの設計の策定を進めている。とくに，多様な日常会話をできるだけ網羅して記録するために，現代日本人がどのような種類の会話をどの程度行っているのかを調査し，それに立脚してコーパスを設計することとした。

本稿では，現在進行中の調査の概要について報告した上で，現段階での中間結果に基づき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，会話収集に当たってどのような調査対象や会話状況を選定すればよいかの指針を得る。

2 会話行動調査

言語社会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。調査の概要を以下に示す。調査の詳細については小磯ほか (2015) を参照されたい。

■調査目的 日常会話の多様性を明らかにし，それに立脚して様々な種類の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 日の日常会話行動の種類や時間などを調査する。

■調査期間 2014 年 11 月 1 日～2015 年 2 月末（予定）

■調査対象日・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

■調査対象 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）。調査協力者（以下，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。調査目的に記した通り，本調査は会話行動の「実態」を明らかにするものではなく，会話行動の多様性を捉えることを目的とする。そのため，家庭での会話行動の多様性の把握のために，家族と同居している人を優先して募集した。

■調査項目 (A) 協力者の属性（性別・年代・職業・世帯員数・居住都道府県），(B) 会話行動に関する調査項目（表 1 の 8 項目）。(B) のうち，会話概要とその他を除く 6 項目を分析項目とする。形式・長さ・相手人数は会話の属性，時間帯・場所・活動は会話状況の属性である。

■調査手続き

- 調査の手引きや調査票（1 日 1 冊，計 3 冊）などの調査資料一式を協力者に事前に郵送した。
- 郵送物が届いてから 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）を選択して調査した。会話の多様性把握のため，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。
- 調査日は調査票を携帯し，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，表 1 に示す 8

表 1　会話行動に関する調査項目

項目	説明
会話概要	会話の内容の簡単な説明【自由記述式】
形式	会話のタイプ（雑談, 用談・相談, 会議・会合, 授業・レッスン・講演）【単数】
長さ	会話の長さ（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）【単数】
相手人数	会話相手との関係（家族, 親戚, 先生生徒, 仕事学業関係, 公共商業関係, 友人知人, 顔見知り・見知らぬ人）と関係ごとの人数【数値入力式】
時間帯	会話が行われた時間帯（午前, 午後, 夜）【単数選択式（以下，単数）】
場所	会話をした場所（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）【単数】
活動	何をしているときに会話をしたか（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）【単数】
その他	該当する項目を選択（電話・ネットでの会話，外国語を交えた会話，外国人を含む会話）【複数選択式】

項目に回答した。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。

■謝礼　3 日間の調査に対し 6000 円。

3　アソシーエーション分析

3.1　アソシエーション分析とは

アソシエーション分析は，データマイニング手法の一種で，コンビニや百貨店などの購買データから関連のある商品の組み合わせを発見するなどの目的で，近年広く用いられている。本研究では，調査項目間の関連をアソシエーション分析で抽出し，ある項目（「会話長さが 5 分未満」など）と関連の深い他の項目を求めることで，その種の会話を収集するにはどのような調査対象や状況を選定すればよいかの指針とする。

3.2　データ

2015 年 1 月 19 日現在の有効回答 200 名分（計 600 日，7559 会話）を対象に分析を行う。200 名の内訳を表 2・3 に示す。

表 2　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	21	23	23	22	23
男性	10	20	16	20	22

表 3　調査対象：職業の内訳

会社員・役員・公務員・専門職	84	自営業	8
パート・アルバイト	29	学生	21
無職・定年退職者	16	その他	11
専業主婦	31		

分析に際し，以下の通り項目の値を一部併合した。

職業　会社員等（会社員・役員・公務員・専門職 + 自営業），パート・アルバイト，学生，専業主婦，その他（無職・定年退職者 + その他）

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，4〜9 人，10 人以上（全ての関係性の相手人数を合計して左記の通り分類）

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

以上で述べた協力者の属性 3 項目（性別 2 種，年代 5 種，職業 5 種），会話の属性 3 項目（形式 4 種，長さ 4 種，相手人数 5 種），会話状況の属性 3 項目（時間帯 3 種，場所 6 種，活動 6 種）をアソシエーション分析で用いた。各項目の絶対頻度を図 1 に示す。

3.3　方法

相関ルールの抽出手法として広く用いられている Apriori アルゴリズム (Agrawal and Srikant, 1994) を用い，R 言語の `arules` パッケージ中の `apriori` 関数を用いて実行した。項目数が多く，またなるべく多様なルールを抽出するため，支持度の最小値は 0.01，確信度の最小値は 0.1 に設定した。また，ルールの長さの最大値は 5 とした。

抽出するルールは，コーパス設計に反映させる可能性のある，形式・長さ・相手人数が結論部に現れるものに限定し，これらの項目がいくつかの特徴的な値を取

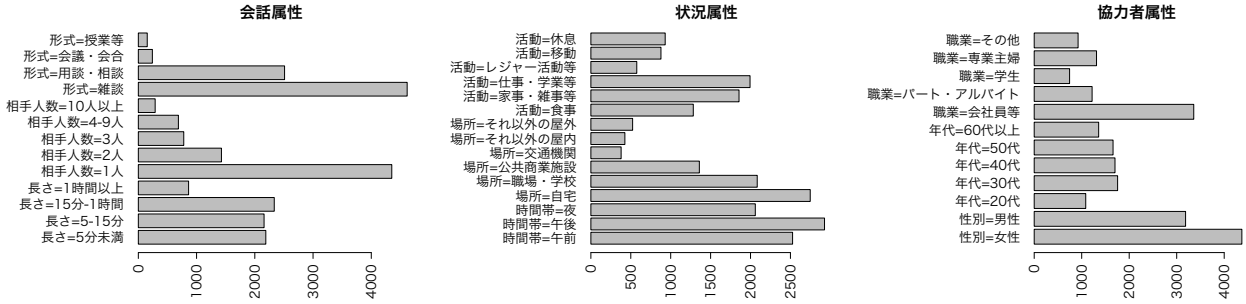

図 1　各アイテムの絶対頻度

る場合に焦点を当てて，その条件部を検討した。抽出されるルール数が非常に多くなるため，affinity 類似度 (Aggarwal et al., 2002) を尺度とするクラスター分析 (Ward 法) によりルールをまとめ上げ，クラスターごとに条件部を要約した。なお，リフトが 1 以下のルールはクラスター分析からは除外した。

3.4　結果

形式・長さ・相手人数を結論部に含む相関ルールがそれぞれ 3425, 3201, 3045 個抽出された。以下，個別に検討する。

3.4.1　形式

3425 個のルールのうち，形式が「雑談」「用談・相談」「会議・会合」「授業等」に関するものはそれぞれ，2226, 1166, 33, 0 個であった。会議・会合は収集するにはむしろ特殊な場面であるため，ここでは「雑談」と「用談・相談」に焦点を当てる。

「形式=雑談」を結論部に持つルールのクラスター分析の結果，4 つのクラスターにまとめることができた。各クラスターの条件部を検討した結果，以下のような解釈が可能であった。解釈には確信度の高いルールをおもに用いた。

1. 非有職者（専業主婦や無職・定年退職者等）の自宅での食事中
2. 会社員等の自宅や職場での食事中
3. さまざまな場所・活動中の学生
4. 交通機関や屋外における移動中

次に，「形式=用談・相談」を結論部に持つルールのクラスター分析の結果，2 つのクラスターにまとめることができた。

1. 女性の公共商業施設における家事・雑事（買い物等）中
2. 男性会社員等の職場における仕事中

3.4.2　長さ

3201 個のルールのうち，長さが「5 分未満」「5〜15 分」「15 分〜1 時間」「1 時間以上」に関するものはそれぞれ，1030, 965, 1052, 154 個であった。ここでは，とくに短い/長い「5 分未満」と「1 時間以上」に焦点を当てる。

「長さ=5 分未満」を結論部に持つルールのクラスター分析の結果，3 つのクラスターにまとめることができた。

1. 女性の公共商業施設における雑事（買い物等）中の用談・相談
2. 学生の学校や自宅における学業・雑事中の雑談や用談・相談
3. 会社員やアルバイト等の職場や公共商業施設における仕事や雑事中の用談・相談

「長さ=1 時間以上」を結論部に持つルールのクラスター分析の結果，3 つのクラスターにまとめることができた。

1. 授業や会議・会合
2. 自宅や公共商業施設における食事や休息中の雑談
3. 女子学生の雑談

3.4.3　相手人数

3045 個のルールのうち，相手人数が「1 人」「2 人」「3 人」「4〜9 人」「10 人以上」に関するものはそれぞれ，2294, 482, 166, 93, 10 個であった。会話相手が 1 人（2 人会話）は大多数を占め，収集するに特段工夫が必要とは思われないため，ここでは，2 人会話を除いてとくに相手人数が少ない/多い「2 人」と「10 人以上」に焦点を当てる。

「相手人数=2 人」を結論部に持つルールのクラス

ター分析の結果，2 つのクラスターにまとめることができた。

1. 女性の自宅での雑談
2. 会社員等の自宅での食事中の雑談

「相手人数=10 人以上」を結論部に持つルールは 10 個しかなく，これらの条件部の特徴は以下であった。

1. 職場や学校での 1 時間以上の会話

4 おわりに

本稿では，多様な会話行動を納めた日本語日常会話コーパスを設計するために現在実施している現代日本人の会話行動に関する調査の概要と，現段階での中間結果について報告した。

調査項目は大きく，調査協力者（会話者）の属性（性別や年代，職業など），会話の属性（会話の形式や長さ，人数など），会話の状況（会話が行われた時間帯や場所，その時の活動の種類など）の 3 つに分かれる。このうち会話者の属性については，こうした調査だけでなく，国勢調査の性別・年層・職業別の人口比なども勘案した上で，会話コーパスの設計を立案することになる。

そこで本稿では，会話の形式や長さ，人数といった会話の属性を中心としたコーパス設計の指針を得るべく，その一つの方法として，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た。その結果，

1. 食事中の雑談や学生の雑談に加え，移動中の雑談や買い物中・仕事中の用談・相談もあること
2. 5 分未満の短い会話の大半がちょっとした用談・相談であるのに対して，1 時間以上の長い会話には授業や会議・会合以外に食事中や女子学生の雑談もあること
3. 3 人会話は自宅での雑談が多く，10 人以上の大人数での会話は 1 時間以上に及ぶものに見られること

などが分かった。今後，これらの結果を詳細に吟味するとともに，さらに分析を進め，コーパス設計に反映させたい。

参考文献

Aggarwal, Charu C., Cecilia Procopiuc, and Philip S. Yu (2002). “Finding localized associations in market basket data.” *IEEE Transactions on Knowledge and Data Engineering*, 14:1, pp. 51–62.

Agrawal, Rakesh, and Ramakrishnan Srikant (1994). “Fast algorithms for mining association rules in large databases.” *Proceedings of the 20th International Conference on Very Large Data Bases* pp. 487–499. Santiago, Chile.

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」 『第 7 回コーパス日本語学ワークショップ予稿集』.