

初対面会話コーパスの構築と分析: 話の盛り上がりと対話時間の関係性

千野愛実花 寺岡丈博
拓殖大学大学院工学研究科

{25d301@st, tteraoka@cs}.takushoku-u.ac.jp

概要

本研究では初対面における日本語母語話者同士および母語話者と学習者の会話コーパスを構築した。本コーパスは、時間的制約を設けず、実際の初対面会話に近い自然な会話データの収集を目的としている。音声・動画・テキストデータに加え、各発話に対して話者自身による感情と盛り上がりのスコアが付与されている。本論文では、コーパスの概要として収集方法と分析結果の一部を示す。ここでは、話者間における盛り上がりの一致が対話時間と関係していることと、母語話者と学習者では対話のペースや語彙の多様さに違いがあることを明らかにした。

1 背景と目的

日本語の会話コーパスは、言語学や情報工学など、幅広い分野において活用されている。これらのコーパスは、会話分析や人間らしい対話システムの構築など、多様な目的のもと設計されている。代表的な例として、日本語日常会話コーパス (CEJC) [1] は、人間同士の自然な会話の特徴を明らかにすることを目的として構築されたものである。一方、Hazumi[2] や Jiang ら [3] の研究では、対話システムに応用するために、発話だけでなく、表情や視線などのマルチモーダル情報を含むデータが収集されている。これらの既存の会話コーパスは、時間的制約がある中で収録された雑談対話がほとんどである。

本研究では、時間的制約がない初対面会話のデータを収集することで、より人間同士の実際の会話に近いコーパスの構築を目指す。さらに、日本語母語話者同士の会話に加え、母語話者と日本語学習者の初対面会話も含むことで、日本語学習者が母語話者と初対面でやり取りする際の指針になることが期待できる。また、本コーパスには、収録した会話データだけでなく、話者が自身と相手の全発話に対して

付与した感情と盛り上がりのラベルが含まれている。そのため、話者の感情や話の盛り上がりを考慮した会話の流れや終わり方を分析できるようになる。したがって、会話分析研究と人間らしい対話システムの実現の両面に貢献することが期待できる。

2 関連研究

これまで様々な会話コーパスが構築されている。会話コーパスは、音声や動画、テキストデータから構成されているコーパスと、発話に加えて、対話者の動作や視線といったマルチモーダルな情報が収録されているコーパスの大きく2種類に分類できる。

前者は、TDU-Kao, CEJC および BTSJ が挙げられる。TDU-Kao [4] は、初対面の70歳以上の女性が10分程度対話する様子を収録している。このコーパスは、高齢者が新たに友人関係を構築するための支援方法を研究するため構築されている。CEJC は、20代から60代の日本語母語話者の対話データが収録されており、自然な会話を収録するために、研究者を介さず話者自身がタイミングを決め収録を実施している。BTSJ [5] は、1対話20分程度の対話が収録されており、発話の重なりや沈黙などを含んでいる。また、話者の社会的属性や対話相手との関係性の分析を重視し、それらが統制されている。

一方、後者のマルチモーダル対話に焦点を当てている研究として、Hazumi や Jiang らの研究が挙げられる。Hazumi は、人とエージェントの対話が収録されているコーパスである。このコーパスは、システム発話の開始から次のシステム発話の開始までを1回とし、本人と第三者による7段階の感情スコアが付与されている。結果として、第三者と本人の感情スコアの相関は中程度であることが明らかにされている [6]。Jiang らは、センサ信号や対話者の主観的感情スコアを含む日本語のマルチモーダル人間対話データセットを構築し、生理信号と主観的評価の

表 1: コーパスの概要

	日本語母語話者		母語話者と学習者
収録期間	2024 年収録	2025 年収録	2025 年収録
内容	2 者間の初対面会話 (雑談)		
合計対話数	20	23	31
合計対話時間	11.5 h	22.9 h	29.5 h
参加者	母語話者 14 名 (男性 9 名, 女性 5 名)		母語話者 26 名 (男性 13 名, 女性 13 名) 学習者 9 名 (男性 6 名, 女性 3 名)
	大学生		大学生, 大学院生
	男性同士 5 対話 - 異性 15 対話	男性同士 7 対話 女性同士 5 対話 異性 11 対話	男性同士 16 対話 女性同士 2 対話 異性 3 対話
アンケート	コミュニケーション能力, TIPI-J, 日本語能力試験 (JLPT) レベル (学習者のみ)		
アノテーション	発話ごとの感情スコアと盛り上がり		
言語	日本語		

表 2: コーパスの内容比較

Corpus	初対面	学習者	感情	盛り上がり	時間無制限
TDU-Kao	○	-	-	-	-
CEJC	-	-	-	-	○
BTSJ	○	○	-	-	-
Hazumi	○	-	○	-	-
Jiang+, 2024	○	-	○	-	-
本コーパス	○	○	○	○	○



(a) 俯瞰カメラ (b) 正面カメラ
図 1: 収録の様子

関係を分析している。このデータセットでは、主観的感情スコアを 10 段階評価とし、1 秒あたり 4 回の頻度で付与している。分析の結果、ユーザの感情スコアに関連する生理信号が明らかにされている。

日本語の会話コーパスはいくつか存在するが、時間を決めずに初対面で会話する様子を収録したコーパスは存在しない。また、これらのコーパスには、話者自身と対話相手の全発話に対して感情と盛り上がりのラベルが付与されてるものはない。本コーパスと他の会話コーパスとの比較を表 2 に示す。

3 データ収集

本研究は、日本語母語話者同士と、日本語母語話者と学習者の 2 種類の初対面对話を収録している。これらのデータ収集には、合計 45 名 (男性 27 名, 女性 18 名) が参加した。そのうち、日本語母語話者は、合計 36 名 (男性 21 名, 女性 15 名)、日本語学習者は、合計 9 名 (男性 6 名, 女性 3 名) である。

また、学習者には、日本語能力試験 JLPT における N1 レベル 4 名 (男性 4 名)、N2 レベル 4 名 (男性 2 名, 女性 2 名)、N3 レベル 1 名 (女性 1 名) が含まれる。参加者は大学生と大学院生であり、1 人最大 5 対話まで収録している。基本的にどの対話も終了時間が決められていないが、個人によっては授業時間などが理由で対話を終了したケースがあった。参加者は事前に収録と公開について承諾しており、当日内に感情と盛り上がりに関する主観的評価を行なっている。なお、本収録は当機関の倫理委員会により承認されている。表 1 は、コーパスの概要を示す。

3.1 収録の環境

映像を撮る 3 台のカメラと音声を取るマイクを用いて、音声、動画及びテキストデータを収集した。全体の映像を撮るカメラ (俯瞰カメラ) は Sony の「FDR-AX45A」を使用、参加者の正面を撮るカメラ (正面カメラ) は「VLOGCAM ZV-1F」を使用、参加者の音声を取るマイクは「ECM-W3」を使用した。

俯瞰カメラは話者の上半身が映る画角で撮影を行った。また、マイクによる音声の収録もこのカメラで行った。正面カメラは、話者が中心に映る画角を撮影ごとに調節した (図 1)。

3.2 アンケート

収録開始前に参加者はアンケートに回答した。アンケートの項目には、人口統計情報 (年齢と性別)、コミュニケーション能力および性格を評価する BigFive[7][8] が含まれている。上記に加え、日本語学習者には日本語能力試験 JLPT のレベルも尋ねており、参加者全員が「初対面の会話は得意である」、「初対面の会話は苦手ではない」という 2 質問

表 3: 対話時間

		対話数	対話時間
母語話者同士	合計	43	34.4 h
	平均 (1 対話あたり)	-	48.1 min
母語話者と学習者	合計	31	29.5 h
	平均 (1 対話あたり)	-	57.2 min
全体	合計	74	64.0 h
	平均 (1 対話あたり)	-	51.9 min

表 4: 平均対話時間の比較 (初対面のデータのみ)

Corpus	母語話者同士	母語話者と学習者
TDU-Kao	10.3 min	-
BTSJ	17.9 min	18.3 min
Jiang+, 2024	10.0 min	-
本コーパス	48.1 min	57.2 min

に対して「はい」もしくは「いいえ」で回答した。なお、ここではコミュニケーション能力を次のように定義する。両方の質問に「はい」と回答した参加者は「コミュニケーションが得意」、最初の質問に「いいえ」、2つ目の質問に「はい」と回答した参加者は「コミュニケーションが得意でもなければ苦手でもない (どちらでもない)」、両方の質問に「いいえ」と回答した参加者は「コミュニケーションが苦手」と分類した。また、BigFive のアンケートには、TIPI-J [9] を利用した。

3.3 アノテーション

収録した当日中に、参加者は感情と盛り上がりに関してアノテーションを実施した。発話は、「楽々ぎじろー君 Pro¹⁾」を用いてオフラインで書き起こし、人手で修正を加えながら話者を識別した。感情は、「-1」(Negative), “0” (Neutral), “1” (Positive) の3段階、盛り上がりは、該当箇所にラベルが付与されている。ここでは、話者自身の発話だけではなく、相手の発話にもラベルが付与されている。

4 分析結果と考察

4.1 対話時間

本コーパスの収録対話数と対話時間を表 3 に示す。日本語母語話者同士は、総数 43 対話、総対話時間は約 34.4 時間であり、1 対話における平均対話時間は約 48 分であった。一方、母語話者と学習者は、総数 31 対話、総対話時間は約 29.5 時間であり、1 対話における平均対話時間は約 57 分であった。

1) <https://tokyo-properties.co.jp/gijiro/>

表 5: コミュニケーション能力別の対話時間

コミュニケーション能力話者 A	話者 B	母語話者同士		母語話者と学習者	
		対話数	平均対話時間	対話数	平均対話時間
得意	得意	3	65.9 min	4	67.7 min
得意	どちらでもない	10	46.2 min	11	52.1 min
得意	苦手	3	70.2 min	6	71.6 min
どちらでもない	どちらでもない	13	47.0 min	2	34.0 min
どちらでもない	苦手	13	40.3 min	6	52.8 min
苦手	苦手	1	61.6 min	2	57.1 min

表 6: 盛り上がりと対話時間の関係。アスタリスク (*) は不一致に対する統計的有意差を示す ($p < 0.05$)。)

	盛り上がり的一致		盛り上がりの不一致	
	対話数	平均対話時間	対話数	平均対話時間
母語話者同士	26	0:53:04	17	0:40:23
母語話者と学習者	20	1:01:23	11	0:49:32
全体	46	0:56:41*	28	0:43:59

また、表 4 に他のコーパスとの対話時間の比較を示す。ここでは、初対面の会話のみ抜粋した平均対話時間を記載している。表 4 より、本コーパスが他のコーパスに比べて、長い対話を収録していることがわかる。これは対話時間が決められていなかったからであり、日常に近い雑談が収録できているといえる。そのため、本コーパスでは対話の始めから盛り上がる過程、さらには終わり方まで分析することが可能である。

コミュニケーション能力別の平均対話時間を表 5 に示す。表 5 より、対話時間が最も長い話者のペアは、「得意」と「苦手」であり、平均対話時間は1時間を超えている。これは、母語話者同士および母語話者と学習者において、同様の結果となっている。一方、対話時間が最も短い話者のペアは、母語話者同士は「どちらでもない」と「苦手」であり、平均対話時間は約 40 分であった。母語話者と学習者は、「どちらでもない」同士が最も短く、平均対話時間は約 34 分であった。さらに、対話時間が長い話者のペアに着目すると、母語話者同士および母語話者と学習者のどちらも、2 番目は「得意」同士、3 番目は「苦手」同士である。これらの結果より、コミュニケーション能力と対話時間はほとんど関係していないことがわかる。

盛り上がりと対話時間の関係については、話者がお互い付与した盛り上がりのラベルが一致した対話と、全く一致しなかった対話の平均対話時間を比較した。その結果を表 6 に示す。前者の平均対話時間の方が長いとともに統計的に有意であることを t 検定により確認した。

上記の結果から、対話時間が長くなる要因は、コ

表 7: 共通の話者 (A) と学習者またはコミュニケーション能力が同じ母語話者 (B) との対話に関する統計情報。アスタリスク (*) は母語話者と学習者間の統計的有意差を示す (p<0.05)。

	対話数	対話全体			1分あたり		話者 A			話者 B			
		対話時間	発話数	形態素数	発話数	形態素数	形態素数	異なり	多様性	形態素数	異なり	多様性	
母語話者同士 (話者 A と話者 B)	合計	5	4.3 h	8,860	55,048	—	—	26,079	2,158	—	28,969	2,542	—
	平均	1	51.9 min	1,772	11,009	32.7*	206.2*	4,894	638	0.769	6,114	897	0.785
母語話者と学習者 (話者 A と話者 B)	合計	5	4.9 h	6,518	53,578	—	—	23,146	1,922	—	30,413	2,750	—
	平均	1	58.8 min	1,304	10,225	20.2	151.6	4,195	566	0.777	6,025	786	0.772

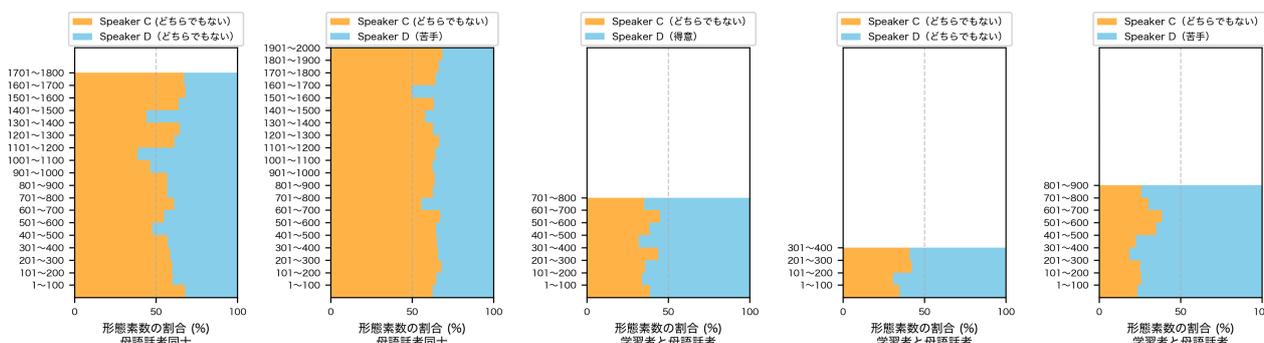


図 2: 100 発話ごとの形態素数の割合

コミュニケーション能力ではなく、盛り上がりの一致が関係していると考えられる。

4.2 母語話者と学習者の比較

表 7 に、母語話者同士および母語話者と学習者の対話に関する統計情報を示す。このとき、話者 A は共通の母語話者であり、話者 B はそれぞれ母語話者と学習者であるが、同じコミュニケーション能力である。ここでは、各々 5 対話に対し分析した。形態素数は、McCab[10] を使用し、辞書は mecab-ipadic-NEologd[11] を用いた。ここでの多様性とは、語彙的多様性を示し、日本語日常対話コーパス[12]と同様、Herdan の C[13] を用いて算出した。

表 7 より、平均対話時間は、母語話者同士に比べ、学習者との対話では 7 分程度長いが、平均発話数は少ない。また、1 分あたりの発話数と形態素数を比較すると、学習者との対話に比べ、母語話者同士の方が多く、その差は t 検定により統計的有意差が認められた。これらの結果から、母語話者同士の対話は学習者よりもペースが速いことがわかる。

語彙多様性に着目すると、母語話者の多様性スコアの平均は 0.785 であり、学習者の平均が 0.772 であることから、母語話者が学習者に比べ、対話中に利用する語彙が豊富であるといえる。さらに、共通の母語話者 A に着目すると、母語話者同士での平均スコアは、0.769 に対し、学習者との対話では、0.777 とわずかに増えていることがわかる。この結果は、母語話者が同じ母語話者との対話に比べ、学

習者が理解できるような簡単な日本語で言い換えることが加わっていたからである。

図 2 に、共通の話者 C (母語話者あるいは学習者) とコミュニケーション能力が異なる話者 D の対話において、100 発話ごとの形態素数の割合を示す。左側 2 つのパネルは、母語話者同士の形態素数の割合であり、右側 3 つは、学習者と母語話者の形態素数の割合である。この図より、母語話者同士の話者 D を比較すると、コミュニケーションが「苦手」な話者の方が、「どちらでもない」に比べ、わずかに形態素数が減っていることがわかる。しかし、学習者との対話では、一貫して話者 D (母語話者) の形態素数が多い。このことから、母語話者と学習者との対話では、コミュニケーション能力に関わらず、母語話者が多く話すことにより、対話が継続していることがわかる。

5 まとめ

本研究では、母語話者同士と母語話者と学習者の 2 種類の初対面会話を収録し、コーパスを構築した。ここでは、対話の終了を参加者の自由としているため、1 対話あたりの対話時間が長く、より自然な対話が収集されていることを確認した。

分析の結果、盛り上がりの一致箇所が多い場合、対話時間が長くなる傾向が見られ、対話の継続と盛り上がりの関連が示唆された。また、母語話者と学習者では、対話のペースや、対話中に使用する語彙の多様さに違いがあることが明らかとなった。

謝辞

本研究は JSPS 科研費 JP22K00646 と JP25K04218 の助成を受けたものです。

参考文献

- [1] Hanae Koiso, Haruka Amatani, Yasuharu Den, Yuriko Iseki, Yuichi Ishimoto, Wakako Kashino, Yoshiko Kawabata, Ken'ya Nishikawa, Yayoi Tanaka, Yasuyuki Usuda, and Yuka Watanabe. Design and Evaluation of the Corpus of Everyday Japanese Conversation. In **Proceedings of the Thirteenth Language Resources and Evaluation Conference**, pp. 5587–5594, Marseille, France, jun 2022. European Language Resources Association.
- [2] 駒谷和範, 岡田将吾. マルチモーダル対話コーパス Hazumi. 自然言語処理, Vol. 29, No. 4, pp. 1322–1329, 2022.
- [3] Jingjing Jiang, Ao Guo, and Ryuichiro Higashinaka. Estimating the Emotional Valence of Interlocutors Using Heterogeneous Sensors in Human-Human Dialogue. In **Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue**, pp. 718–727, Kyoto, Japan, sep 2024. Association for Computational Linguistics.
- [4] 武川直樹, 西岡菜月, 山田晴奈, 楊井一彦, 高柳直人, 平石牧子, 徳永弘子. 自己開示への返答による交友関係構築のケーススタディ-高齢女性による初対面会話を対象に-. 電子電子情報通信学会論文誌 A 基礎・境界, Vol. J106-A, No. 3, pp. 125–135, 2023.
- [5] 宇佐美まゆみ. BTSJ1000 人日本語自然会話コーパス. 科研基盤研究 (A) 「語用論的分析のための日本語 1000 人自然会話コーパスの構築とその多角的研究」(研究代表者: 宇佐美まゆみ) 及び, 国立国語研究所機関拠点型基幹研究プロジェクト「日本語学習者のコミュニケーションの多角的解明」(2016–2021), 2023.
- [6] Kazunori Komatani, Ryu Takeda, and Shogo Okada. Analyzing Differences in Subjective Annotations by Participants and Third-party Annotators in Multimodal Dialogue Corpus. In **Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue**, pp. 104–113, Prague, Czechia, September 2023. Association for Computational Linguistics.
- [7] Lewis R. Goldberg. An alternative description of personality: The Big-Five factor structure. **Journal of Personality and Social Psychology**, Vol. 59, No. 6, pp. 1216–1229, 1990.
- [8] Robert R. McCrae and Oliver P. John. An Introduction to the Five-Factor Model and its Applications. **Journal of Personality**, Vol. 60, No. 2, pp. 175–215, 06 1992.
- [9] 小塩真司, 阿部晋吾, Pino Cutrone. 日本語版 Ten Item Personality Inventory (TIPI-J) 作成の試み. パーソナリティ研究, Vol. 21, No. 1, pp. 40–52, 2012.
- [10] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying Conditional Random Fields to Japanese Morphological Analysis. In **Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing**, pp. 230–237, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [11] 佐藤敏紀, 橋本泰一, 奥村学. 単語分かち書き辞書 mecab-ipadic-neologd の実装と情報検索における効果的な使用方法の検討. 言語処理学会第 23 回年次大会発表論文集, pp. 875–878, 3 2017.
- [12] 赤間怜奈, 磯部順子, 鈴木潤, 乾健太郎. 日本語日常対話コーパスの構築. 言語処理学会 第 29 回年次大会発表論文集, pp. 108–113, 2023.
- [13] Gustav Herdan. **Type-token mathematics : a textbook of mathematical linguistics**. Janua linguarum. Mouton, 1960.