

# 商談対話における階層的 Dialogue Act スキーマの提案

戸田 隆道<sup>1</sup> 二宮 大空<sup>1</sup> 天谷 航平<sup>1</sup> 野崎 文宏<sup>1</sup> 干飯 啓太<sup>1</sup>

<sup>1</sup> 株式会社 AI Shift

{toda\_takamichi, ninomiya\_hirokata, amagai\_kohei, nozaki\_fumihiko, hoshii\_keita}@cyberagent.co.jp

## 概要

企業の営業活動において商談データの分析は重要であるが、プライバシー保護の観点から実データの共有・活用は困難である。この課題に対し、構造を維持した合成データの生成が期待されているが、商談特有の文脈構造を記述する体系的な枠組みは確立されていない。本研究では、商談対話の構造的特徴を捉えるための階層的 Dialogue Act スキーマを提案する。提案スキーマは、商談の進行段階を表す5種の Phase と、個々の発話行為を表す9種の Act から構成される二層構造を持つ。擬似的な B2B SaaS 商談データに対してアノテーションを行い、スキーマの適用可能性を検証した。提案スキーマは、商談対話の構造的な理解および合成データ生成パイプラインの基盤となることが期待される。

## 1 はじめに

企業の営業活動において、商談データは成約予測モデルの構築や営業教育 AI の開発に不可欠な資源である。しかし、商談の書き起こしデータには顧客の個人情報や企業の機密情報が含まれており、プライバシー保護の観点から外部公開が困難である。

この課題に対し、合成データの生成が有効なアプローチとなりうる。ただし、ゼロから生成した合成データはリアリティに欠け、実用的な分析に耐えない [4]。商談の構造を維持しつつ、文脈を架空のものへ置換するアプローチが必要である。

そのためには、まず商談対話の構造を明示的に理解する必要がある。本研究では、商談対話に特化した階層的 Dialogue Act スキーマを提案する。

## 2 関連研究

対話行為 (Dialogue Act) の研究では、ISO 24617-2 に代表される国際標準が確立されている [1]。しかし、これらは汎用的な機能分類に留まり、商談のような長期的な文脈依存性や進行プロセスを持つ対話

を記述するには粒度が粗いという課題がある。

ドメイン特化型スキーマとしては、ABCD がコールセンター対話を対象に Action State Tracking を導入し [3]、CaSiNo が交渉対話における戦略レベルのアノテーションを提供している [2]。しかし、これらはいずれも B2B 商談を対象としたものではなく、商談特有の段階的進行や役割分担を体系的に表現するスキーマは提案されていない。

一方、営業対話の分析分野では、産業界において Conversation Intelligence としての実用化が進んでいる。代表的な商用サービスとして、Gong<sup>1)</sup> や Chorus.ai<sup>2)</sup> が存在し、営業通話の自動録音・書き起こし・分析を通じて、成約予測や営業コーチングを支援している。Gong らの調査によれば、トップセールスに共通する発話率 (Talk-to-Listen Ratio) や質問頻度などの定量的指標が成約予測に有効であることが示されている<sup>3)</sup>

学術研究においても、成約予測や戦略抽出が主流である。特に近年の Nandakishor らは、GPT-4o により生成した合成データを用いて強化学習でモデルを学習し、プライバシー制約の厳しい営業領域において、合成データが実データの代替として極めて有効であることを実証した [4]。しかし、これらのアプローチは発話長、感情、キーワードといった対話の表層の特徴や応用タスクに焦点を当てており、大規模言語モデルの確率的生成能力に強く依存している。

これに対し本研究は、商談を「Phase (進行段階)」と「Act (発話行為)」の二層構造で捉える階層的スキーマを提案する点に独自性を持つ。対話構造を人間が解釈・制御可能な形式で明示的に定義することで、商談のような定型的な流れを持つ対話において不可欠となる構造的整合性を担保しつつ、高品質かつ説明可能な合成データ生成の基盤を提供する。

1) <https://www.gong.io/>

2) <https://www.chorus.ai/>

3) <https://www.gong.io/blog/talk-to-listen-conversion-ratio/>

### 3 提案スキーマ

#### 3.1 設計方針

提案スキーマは以下の 3 つの設計方針に基づき構成された。

1. 階層構造：商談の大局的な流れ（Phase）と、個々の発話行為（Act）の二重階層構造
2. 実データ駆動：営業理論を参考にしつつ、実際の商談データから帰納的に導出されたスキーマ構成
3. 簡潔性：アノテーション可能な粒度を維持しつつ、過度な細分化を避けることを重視した設計

#### 3.2 Phase 層

商談の進行段階を表す 5 種類の Phase を定義した（表 1）。

表 1 Phase 層の定義

Phase	説明
Opening	商談の開始. 挨拶, 自己紹介, 本日の目的・アジェンダの提示
Situation	顧客状況の把握. 組織体制, 業務フロー, 導入済みツールのヒアリング
Proposal	提案・説明. 会社紹介, 製品説明, デモンストレーション, 価格提示
Objection	異議・質疑対応. 顧客からの質問や懸念への回答
Closing	商談の終結. 次回アポイント設定, まとめ, 御礼

提案スキーマでは、Phase の出現順序は固定されていない。典型的な商談フローは Opening → Situation → Proposal → Objection → Closing の順序を取るが、実際の商談では、Objection が複数回出現する、Situation の後に再度 Situation が来る、一部の Phase が省略される（例：初回商談で Proposal が省略される）など、柔軟な順序や省略が許容される。本スキーマは、このような実務的な多様性を反映し、各 Phase を独立した段階として定義している。

#### 3.3 Act 層

個々の発話行為を表す 9 種の Act を定義する。Phase 横断的に出現する汎用 Act (3 種) と、特定 Phase に紐づく Phase 固有 Act (6 種) に分類する（表 2）。

表 2 Act 層の定義

分類	Act	出現 Phase	説明
汎用	Question	全 Phase	質問する
汎用	Respond	全 Phase	回答・情報開示
汎用	Acknowledge	全 Phase	相槌・挨拶・感謝
固有	Introduce	Opening	自己紹介・目的説明
固有	Present	Proposal	説明・デモ・価格提示
固有	Raise	Objection	懸念・質問の提起
固有	Handle	Objection	懸念への対応
固有	CheckIntent	Closing	導入意思・温度感の確認
固有	ScheduleNext	Closing	次回日程調整

Handle と Respond はともに「回答」を表す Act であるが、その適用範囲が異なる。Handle は Objection Phase に限定され、Raise で提起された懸念や質問に対する回答に特化した Act である。一方、Respond は全 Phase で使用可能な汎用 Act であり、Handle が対象としない一般的な質問への回答や情報開示を表す。例えば、Situation Phase における顧客の質問（Question）に対する回答は Respond として分類されるが、Objection Phase において Raise で提起された懸念への対応は Handle として分類される。

### 4 検証実験

本章では、提案する階層的 Dialogue Act スキーマの妥当性、および LLM を用いた自動アノテーションの実用性を検証するための実験設定について述べる。

#### 4.1 検証仮説

提案スキーマの妥当性を検証するため、以下の 3 つの仮説を設定した。

- 仮説 1（スキーマの客観性）：提案スキーマは個人の主観に依存せず、異なるアノテーター間で高い一致率が得られる客観的な分類体系である。
- 仮説 2（スキーマの実データ整合性）：提案スキーマは、設計時に想定した商談の進行構造と実データ上の Phase-Act の出現パターンが整合的である。

- **仮説 3 (階層構造の有効性)** : Phase 固有 Act は特定の Phase に集中して出現し, Phase 情報を条件とした階層的推論が有効である.
- **仮説 4 (LLM による自動化可能性)** : LLM はプロンプト指示のみで高精度な分類を達成でき, 人手を介さない合成データ生成パイプラインへの組み込みが実用的である.

以下, これらの仮説を検証するための実験設定と結果を示す.

## 4.2 データセット

本研究では, B2B SaaS 商材を扱った模擬商談 (ロールプレイ) データセットを使用した. 実データの利用はプライバシー保護の観点から制約が多く, また本研究の目的である「商談構造の体系化」には, ノイズを制御しつつ典型的な商談フローを網羅したデータが適しているためである.

本データセットの特筆すべき特徴として, 発話長の分散が極めて大きい点が挙げられる. 具体的には, 製品の仕様や価値を伝えるための数分間に及ぶ長文の説明が含まれる一方で, 「画面を共有します」「少々お待ちください」といった数秒程度の短い機能的な確認発話 (Functional utterances) も多数混在している. データセットは 1 件の商談セッションから構成され, 総発話数は 174 ターン (営業担当者と顧客の合計 348 発話) である. このセッションは, 挨拶からクロージングまでの一連の流れを含んでおり, 対話時間は約 50 分である. データは Deepgram の Nova3 による音声認識で書き起こされ, 話者分離も行われた後, フィラー除去などの軽微なクリーニング処理を施している.

## 4.3 アノテーション体制

提案スキーマの客観性と信頼性を評価するため, ドメイン知識レベルの異なる以下の 3 属性・計 4 名 (LLM 含む) によるアノテーションを実施した.

- **スキーマ設計者**: 1 名. 提案スキーマの定義を最も厳密に理解している.
- **ドメインエキスパート**: 2 名. 実際の B2B 営業実務に従事する現役の営業職である. 言語学的な定義よりも, 商談現場における実務的な文脈解釈に基づき判断を行う.
- **大規模言語モデル**: 本実験では Gemini 3 Pro を使用した. 人手によるアノテーションコストの

削減および合成データ生成時の制御可能性を検証するため, プロンプトによる指示のみで分類を行わせた.

### 4.3.1 評価指標

スキーマの信頼性と LLM の性能を測るため, 以下の指標を用いる.

- **アノテーター間一致率**: 人間アノテーター 3 名 (スキーマ設計者 1 名およびドメインエキスパート 2 名) の間で判断がどの程度一致したかを評価するため, Fleiss' Kappa 係数を用いる. これにより, 提案スキーマが個人の主観に依存しない客観的な分類体系であるか検証する.
- **LLM の分類精度**: 人間の合意ラベル (多数決) を正解データとし, LLM の予測精度を Accuracy および F1-score により評価する.

## 4.4 実験結果

### 4.4.1 仮説 1 の検証: アノテーター間一致率

仮説 1 を検証するため, Phase 層および Act 層におけるアノテーター間一致率 (Fleiss' Kappa) を評価した. 結果を表 3 に示す.

表 3 アノテーター間一致率

層	Fleiss' Kappa
Phase	0.949
Act	0.959

両層ともに Fleiss' Kappa 係数が 0.949 (Phase 層) および 0.959 (Act 層) と極めて高い値を示し, 仮説 1 が支持された. 提案スキーマが個人の主観に依存しない客観的な分類体系であることが確認された. 特に Act 層においても高い一致率が得られたことは, 9 種類の Act の定義が実務者にとっても直感的に理解可能であることを示唆している.

### 4.4.2 仮説 2, 3 の検証: Phase × Act の出現頻度

仮説 2 および仮説 3 を検証するため, Phase × Act の出現頻度を分析した. 結果を図 1 に示す.

各 Phase において, 設計時に想定した Act が優勢に出現していた. Opening では Introduce/Acknowledge, Situation では Question/Respond, Proposal では Present, Objection では Raise/Handle, Closing では CheckIntent/ScheduleNext が主に観察され, Phase 定義と実データの整合性が確認された (仮説 2

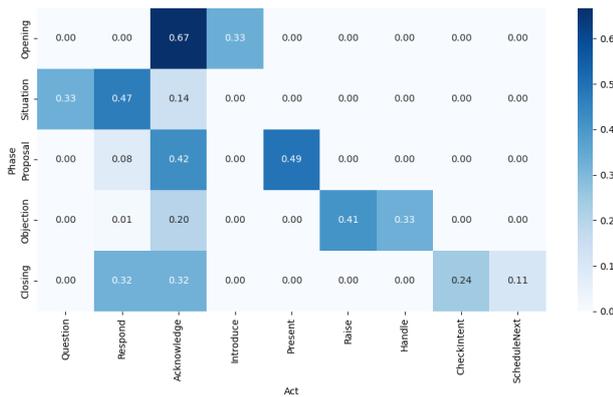


図 1 Phase × Act の正規化出現頻度ヒートマップ. 各 Phase 内で Act の出現割合を正規化して可視化した. 色が濃いほど当該 Phase における Act の出現頻度が高いことを示す.

支持).

また, Phase 固有 Act は特定の Phase に強く偏って出現しており, Phase 情報を条件とした階層的推論の有効性が示唆された (仮説 3 支持).

#### 4.4.3 仮説 4 の検証: LLM による自動分類精度

仮説 4 を検証するため, 人間アノテーター 3 名の多数決ラベルを正解データとし, Gemini 3.0 Pro による自動分類精度を評価した. 結果を表 4 に示す.

層	Accuracy	Macro-F1
Phase	0.954	0.953
Act	0.954	0.800

Phase 層では, Accuracy と Macro-F1 がともに 95% を超え, 仮説 4 が支持された. LLM がプロンプト指示のみで高精度な分類を達成できることが示された. 一方, Act 層では Accuracy が 95.4% と高水準である一方, Macro-F1 は 80.0% に留まった. この乖離は, Act カテゴリ間のクラス不均衡に起因すると考えられる.

## 5 限界と課題

本研究にはいくつかの限界が存在する. 第一に, データの規模と多様性である. 本実験の評価対象は単一の模擬商談セッション (174 ターン, 合計 348 発話) に限定されており, 提案スキーマの汎化性能については, 異なる商材や対話者によるデータを用いた追加検証が必要である. 第二に, 対話環境の差異である. 本研究ではノイズの少ない模擬データを使用した, 実際の商談現場では, 発話の中断, 脱

線, 同時発話などが頻発する. 実務環境特有のノイズに対するスキーマおよびアノテーションの頑健性は未検証である.

## 6 おわりに

本研究では, B2B 商談対話の構造的特徴を捉えるための階層的 Dialogue Act スキーマを提案した. 提案スキーマは, 商談の大局的な進行段階を表す 5 種の Phase と, 局所的な発話行為を表す 9 種の Act から構成される二層構造を持つ. 検証実験により, 設定した 4 つの仮説がすべて支持された.

今後の展望として, 本スキーマを基盤とした「構造制御型合成商談データ生成システム」の実装を計画している. 具体的には, Phase 間の遷移確率と Act の出現パターンを明示的に制御することで, プライバシーを保護しつつ実データの構造的特徴を維持した高品質な合成データの生成を目指す. これにより, データ不足が課題となる成約予測モデルの学習や営業教育 AI の開発において, 実データの代替として活用可能なデータセットの構築が期待される.

## 参考文献

- [1] Harry Bunt, Jan Alexandersson, Jean Carletta, et al. Iso 24617-2: A semantically-based standard for dialogue annotation. In **Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)**, 2012.
- [2] Kushal Chawla, Kushal Wadhwa, Pradyumna Singh, et al. Casino: A corpus of strategic negotiation dialogues. In **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)**, pages 254–267, 2021.
- [3] Qian Chen, Zhuohan Li, Yizhe Zhang, et al. Abcd: A large-scale, diverse, english dialogue dataset for action-based conversation modeling. In **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)**, pages 239–255, 2021.
- [4] V. Nandakishor et al. Learning sales dialogue policies with synthetic data generated by large language models. **arXiv preprint arXiv:2503.23303**, 2025.

## A アノテーション手順の詳細

本付録では、検証実験で実施したアノテーション手順の詳細について述べる。

### A.1 データ前処理

本研究で使用した商談データは、以下の前処理を経てアノテーションに供された。

- **音声認識 (STT)** : Deepgram Nova 3 を使用して音声をテキストに書き起こした。
- **話者分離**: オンライン Meeting プラットフォームの自動話者分離機能により、営業担当者と顧客の発話を自動的に分離した。
- **クリーニング**: フィラー除去などの軽微なクリーニング処理を実施した。

これらの前処理により、話者情報が付与された書き起こしテキストが生成され、各発話に対して Phase および Act ラベルのアノテーションが可能となった。

### A.2 LLM による自動アノテーション

本研究では、LLM による自動アノテーションの実用性を検証するため、Gemini 3.0 Pro を使用した。

#### A.2.1 プロンプト設計

LLM への指示プロンプトは以下の通りである。  
商談の DialogAct をアノテーションしています。商談構造は Phase → Act という 2 層構造になっていると想定しています。書き起こしに以下のラベルをアノテーションしてください。  
**ラベル**  
商談の構造は Phase → Act という 2 層構造になっているという想定  
**Phase 層** [Phase 層の定義を記載]  
**Act 層** [Act 層の定義を記載]  
**書き起こし** [対象となる書き起こしテキストを記載]

#### A.2.2 コンテキスト情報

LLM による分類においては、以下のコンテキスト情報を提供した。

- **発話の範囲**: 各発話を分類する際、該当発言のみではなく、それ以前の発言も含めた対話履歴をコンテキストとして提供した。これにより、LLM は対話の文脈を理解した上で Phase および Act を判定できる。
- **話者情報**: 音声認識結果には話者分離が行われ

ており、各発話に営業担当者または顧客の話者情報が付与されている。この話者情報も LLM への入力に含めた。

- **時系列情報**: 発話の出現順序を保持し、時系列的な対話の流れを LLM が把握できるようにした。

これらのコンテキスト情報により、LLM は単一の発話を独立に分類するのではなく、対話全体の文脈を考慮した分類を実施できる。特に Phase 層の分類においては、商談の進行段階を正確に把握するために、過去の発話履歴を含めたコンテキストが重要である。