

Dialogue Act 情報を付与したメールコーパス

上田 良寛 谷口 友紀 三浦 康秀 大熊 智子

富士ゼロックス株式会社

{ueda.yoshihiro, Taniguchi.Tomoki, yasuhide.miura,
ohkuma.tomoko}@fujixerox.co.jp

1 はじめに

メールはビジネスにおける主要なコミュニケーションツールであり、メールデータの中には様々な知識が蓄積されている [1]。ただその知識は整理されて存在するのではなく、メッセージのやりとりの連鎖であるスレッド内に分散されている。個々のメールメッセージはそれぞれ役割、例えば、「依頼」、「質問」、「回答」、「情報提供」などを持っており、役割を通してメッセージに分散されている知識を結びつけることができる。

ここでいう「役割」は、一般には Dialogue Act と呼ばれており、Speech Act を対話に発展させたものである。ISO でも標準化がすすめられている [2, 3]。

メールコミュニケーションの分析に Dialogue Act は有力なツールとなると考えている。そのような研究には、メールに Dialogue Act 情報が付与されたコーパスが不可欠である。DailyDialog [4] は Dialogue Act 情報が付与されているものの、英語学習者用にサンプル対話 (口頭での会話) を集めたもので目的にそぐわない。QA に特化した Dialogue Act と言える Intent が付与された MSDialog [5] は、サポートフォーラム (QA 掲示板) のコーパスである。メールに対してアノテーションを付与したものとして、W3C コーパス [6] があるが、ここで付与されている情報は Speech Act であり、対話に適した Dialogue Act と異なる。Cohen らの研究 [7] もメールを対象にしており、“email speech act” という分類体系を使っているが、応用は個別のメールの分類にとどまっている。

このため我々は、オフィスメールに対して、Dialogue Act 情報を付与したコーパスを作成することにした。

本稿では、まずビジネスにおけるメールの利用状況調査に関して報告し、データセット構築に関して説明し、得られたデータセットの分析を行う。また、機械学習の学習データとして利用した場合の評価を行う。

2 メールの利用状況調査

オフィスにおけるコミュニケーションの実態を知るため、ネットアンケートを行った。1000 名の回答を得、そのうちの 475 名がオフィスワーカーであった。

主要なコミュニケーションツールについて質問した (重複回答を許す) 結果、メールが 443 名で、電話の 396 名を抑えトップになっていることがわかった。また、別の質問でメールの使用状況を問うたところ、343 名が「常に使う」、113 名が「時々使う」と答えた。

メールを常にまたは時々使う 456 名のうち、376 名 (82.5%) がメールでの依頼があると答えている。依頼された業務を行う上で、他の人に必要な知識を問い合わせ、有益なメールを転送してもらった人が 302 名 (66.2%)、直接メール返信でもらった人が 316 名 (69.3%) いた。一方、自分の過去メールから探したという人が 393 名 (86.2%) であった。

この結果は多くの知識がメールを媒介して交換され、蓄積されていることを示している。

3 コーパスの作成

3.1 メールの収集

ある企業で働くオフィスワーカー 11 名に同意を得てメールの収集を行なった。これまでに収集したメールメッセージ 7600 通のうち、スレッド (ある話題に関して、利用者間で取り交わされた一連のメッセージの集合) に属するメッセージ数の平均値は 2.41、中央値は 2 であった。メッセージあたりの平均文字数は 1057、中央値は 402 であった。メッセージ当たりの情報量が多い一方で、一連のスレッド中のメッセージ数 (ターン数) が少ないというメール固有の特徴が見てとれる。

3.2 Dialogue Act アノテーション

このようにして集めたメールメッセージに対して、Dialogue Act 情報のアノテーションを行った。アノテーションする内容は Dialogue Act Type だけであるが、この際に、その判断を行なったメッセージの断片 20 文字程度を「エビデンス」として残してもらった。

日本語学を学んだアノテーター 3 名に、最初は同じものに対してアノテーションしてもらい、共通点と差異を見て、マニュアルを精緻化していった。

Dialogue Act 分類 ここで用いる Dialogue Act Type は ISO 分類 [2] から絞った。図 1 は、ISO 分類の上に我々の分類範囲を示し、その上に白抜き文字で名前をつけたものである。ISO 分類との違いを表 1 に示す。

このアノテーションは、前の投稿を参照する必要がある。例えば、Answer とそれ以外の Inform は、前の投稿が Question であるかによって決まる。同意 (「それでいきましょう」) は、前の投稿が Request または Suggest の場合は Commissive (= そのように進めます)、前の投稿が Commissive で何らかの価値提供の

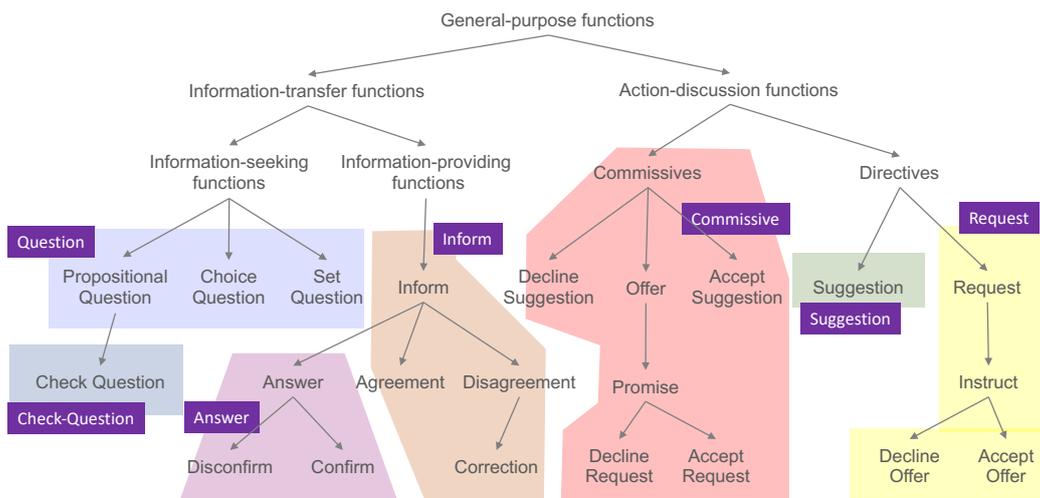


図 1: メールコーパスで用いる Dialogue Act Type (白抜き文字) (背景は ISO の分類)

表 1: Dialogue Act Type の分類 (ISO 分類との差異)

ISO 分類	我々の分類との対応
Directives	受け手の行動に影響を与えるもの。大きく話し手が行って欲しい依頼 Request と、受け手の利益に繋がる提案 Suggestion に分ける
Commissives	話し手が行う行動に関する言明。自分から発する約束だけでなく、Request/Suggestion に対する受諾/拒絶を含み、全体をそのまま Commissive とする
Information-providing functions	情報を提供する。Question に対する回答である Answer と、それ以外の Inform に分ける
Information-seeking functions	情報を求める。答えを求める Question と、話し手が答えを持っていてその確認を求める Check-Question に分ける

申し出の場合は Request (= そのように進めてください)、Commissive だが自分がやることの宣言・約束の場合は Inform (= 賛同します) となる。

Dialogue Act を与える単位 Bunt[2] は文よりさらに小さく Dialogue Act の単位を切っているのに対して、我々は、メッセージ単位を基本として Dialogue Act Type を付与することにした。

一つのメッセージが複数の Dialogue Act を含む場合があるが、それらに主となるものと補助的なものの関係がある場合には、主となる Dialogue Act Type を選択するようにした。例えば、現在の仕様を確認しながら要求を出している場合、Check-Question はつけず、Request だけ記載する。

途中で話題やトーンが変わった時のように、複数の Dialogue Act Type がそれぞれ重要である場合、次の 2 つの対応策が考えられる。1) メッセージ全体に複数の Dialogue Act を与える、2) メッセージ本文を Dialogue Act の切れ目で分割し、それぞれに Dialogue Act を与える。本稿の分析では (1) の方法を用いている。

4 Dialogue Act 分析

ここでは、付与された Dialogue Act Type の割合 (Dialogue Act 分布)、ある Dialogue Act Type の応答としてどのような Dialogue Act Type が来るか (応答

の分布)、本来期待される応答がどの程度充足されるか (期待と充足) を示す。また、主要な Dialogue Act に対して、その特徴を示すパターンを分析する。

4.1 Dialogue Act 分布

Dialogue Act Type 別出現頻度を図 2 に示す。メールでは圧倒的に Inform が多く、Request、Commissive が続く。

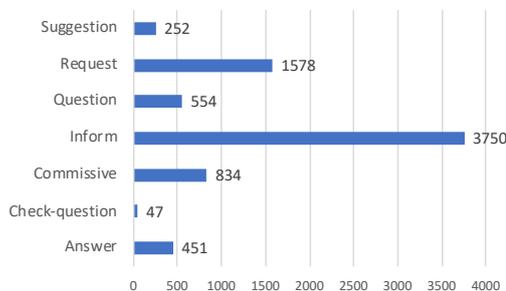


図 2: Dialogue Act Type 別出現頻度

メールと通常の会話の違いを見るため、DailyDialog[4] の 10 種類のトピックのうち Work (仕

事上の会話)のみを抽出して比較する。DailyDialogの Dialogue Act の分類 4 つに揃えて図 3 に示す。

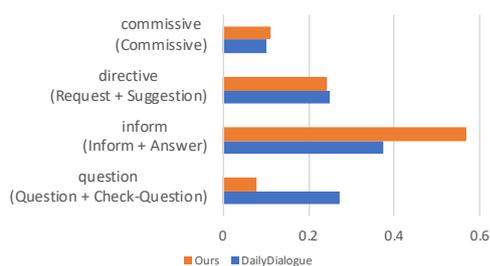


図 3: Dialogue Act Type 割合の DailyDialog との比較

DailyDialog は、question が多く inform が少ないことがわかる。DailyDialog の発話あたりの語数は 14.6 で一連の会話あたりのターン数は 7.9 となっており [4]、3.1 で示したメールより短い単位でターンが切り替わりやりとりが継続する傾向があることがわかる。このため質問がしやすいと考えられる。また、メールメッセージは複数文で構成されるため、例えば Request で状況を確認するための質問などは独立して Question を与えず、Request の一部にしている我々のアノテーション方法もこの差をうむ一因であろう。

4.2 応答の分布/期待される応答とその充足

会話のやりとりで、ある Dialogue Act Type のメッセージがあったときに、それに応答するメッセージにどのような Dialogue Act Type が来るのかを調べた (図 4)。

応答 Dialogue Act → ↓元 Dialogue Act	Question	Check-question	Answer	Request	Suggestion	Commissive	End
Start	0.18	0.01	0.05	0.60	0.07	0.10	0.00
Question	0.06	0.00	0.53	0.08	0.02	0.09	0.21
Check-question	0.02	0.05	0.60	0.05	0.07	0.05	0.17
Answer	0.09	0.02	0.10	0.03	0.03	0.16	0.57
Request	0.08	0.01	0.03	0.21	0.04	0.25	0.39
Suggestion	0.04	0.00	0.01	0.11	0.10	0.20	0.54
Commissive	0.06	0.01	0.03	0.15	0.03	0.15	0.57
Inform	0.04	0.00	0.02	0.10	0.02	0.09	0.73

図 4: 応答の分布 (赤枠は期待される応答)

縦軸は元 Dialogue Act Type、横軸はその応答の Dialogue Act Type である。ただし、Inform はどのような場面でも応答となりうるため、応答 Dialogue Act Type から除いた。会話の最初と最後を示す Start と End を加えた。

ある Dialogue Act が発せられた時に、その応答に期待される Dialogue Act がある。例えば質問 (Question, Check-Question) に対する応答は回答 Answer

であることが望ましい。そのような応答に期待される Dialogue Act Type を赤枠で示す。

Question, Check-question に対しては Answer が多く、これは期待通りである。一方、Request, Suggestion に対しては Commissive (「承知しました」、「～します」など) が、Commissive に対しては Request (「それをお願いします」など) が多く、それぞれ期待に合っているが、それ以上に End (= 会話がそこで終わる場合) が多い。

その理由は、メール以外で (直接対面で) 返したり、返さなくとも依頼事項を実施することで事足りることが多いためという可能性がある。Suggestion に対して Request が低いのは、Suggestion が個別のメッセージであるのに対して、Request では全員への依頼も多いからと考えられる。Commissive が低いのも同様に、全体に対しての実行宣言で、返す必要がない場合が多いと考えられる。

4.3 テキストパターン

それぞれの Dialogue Act Type ごとにその Dialogue Act を特徴付ける表現のパターンを調べた。アノテーション時に「エビデンス」となる部分を抜き出していたが、それをさらに一般化し、「パターン」という項目を設けた。パターンの抽出にあたっては、終助詞の排除、常体/敬体の差の吸収 (全体をどちらかに吸収するのでなくそのパターンで多い方に統一) を行った。

結果の一部を表 2 に示す。Question, Check-Question は文型から疑問形であることがほとんどであり、ここでの分析の対象から外した。

それぞれの上位はその Dialogue Act の役割にあったパターンが抽出されていることがわかる。「～(さ) てください/(さ) させていただきます」が Request と Commissive に現れているが、Request の場合は、「変更させてください (= 変更しますので合わせてください)」など暗に依頼を示すものであった。

少ないパターンにもこのような特徴的な表現が見られる。Request で「ありがとう」が 8 件あったが、Commissive に対して「それで進めてください」の意であった。Commissive の「お願いします」(8 件) は「参加をお願いします (= 参加します)」のような表現で意志を表している。このように表層上の表現では判断できないものも多い。

5 Dialogue Act の自動分類

作成したコーパスは、Dialogue Act の分類モデルを訓練するための教師データとしての活用が期待される。そこで、機械学習モデルによる Dialogue Act の分類性能を評価した。

モデル 談話の構造素性を用いた談話分類手法で用いられる Conditional Random Field (CRF) モデルと Logistic Regression (LR) モデルを評価対象に選定した [8]。メールによる会話は談話の一種であり、メールによって伝達されるメッセージの構造は、談話構造と類似する。そのため、これらの手法により高い分類性

表 2: Dialogue Act を特徴付けるテキストパターン

Request	個数	Suggestion	個数	Commissive	個数
～お願いします/できますか	627	～(の/する) ほうがよい	44	～(し) ます	346
～ください	361	～(し) ましょう	32	～(さ) せてください/～(さ) せていただきます	104
～いただけないでしょうか/～ いただけますか	180	～がいい/～もいい	28	了解しました/了解です	62
～いただけると～	67	いかがでしょう/どうでしょう	26	承りました/承知しました	62
～(さ) せてください/～(さ) せていただきます	64	～可能でしょうか	9	大丈夫です/OK です/問題ありません	30

能が期待される。[8]と同様に分類対象のメッセージに含まれる「本文」と「件名」からそれぞれ作成した n-gram と、その付帯情報値(会話の深さ、文数、単語数、文字数)に加えて、親メッセージの付帯情報を素性として選んだ。さらに、これらの付帯情報を標準化した素性を追加した。

実験設定 [8]に従い、平均 F1 値を評価指標とした。平均 F1 値はスレッド毎に算出した F1 値の平均で定義される。Dialogue Act コーパスを 8:1:1 の割合で分割して、訓練・開発・テストデータを作成した。訓練データでモデルを学習させ、開発データで正則化パラメータを探索した。CRF には L1 正則化を、LR には L2 正則化を適用した。

評価結果と考察 テストデータにおける分類した際の平均 F1 値とラベルごとのデータ数の関係を図 5 に示す。データ量が多い Inform や Request で分類性能が高いことがわかる。また、Question で 27.8pt、Answer で 33.8pt と CRF が LR を大きく上回った。メッセージの系列情報を考慮することが、Dialogue Act を分類する上で重要であることがわかる。

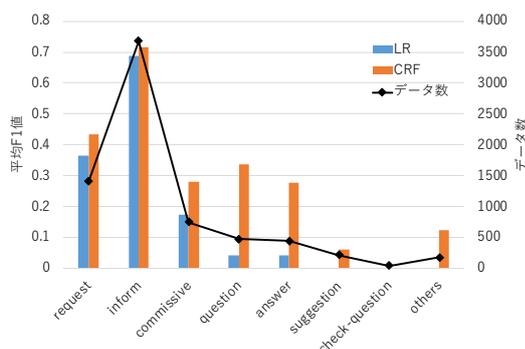


図 5: Dialogue Act の分類性能とデータ量の関係: LR と CRF は左軸、データ量は右軸で表す。

6 おわりに

Dialogue Act 情報を加えたメールコーパスを作成し、分析の結果以下のような知見を得た。

- ・口頭での会話に比べ Inform=情報提供が多く、Question=質問は少ない。(4.1)

- ・オフィスでは対面での情報交換も多いため、必ずしもメールで期待される応答が得られる訳ではない。(4.2)
- ・それぞれの Dialogue Act Type に特有のテキストパターンはあるものの、Dialogue Act が暗黙的に示されているものも多い。(4.3)
- ・このコーパスを学習に用い、Dialogue Act Type の予測に応用可能である。(5)

今後、Dialogue Act Type の予測から、メールの推薦 [1] に対するユーザーの要求を捉えることに発展させたい。

参考文献

- [1] 谷口友紀, 上田良寛, 谷口元樹, 三沢翔太郎, 三浦康秀, 大熊智子. メールデータを活用した社内情報の推薦システム. 言語処理学会, 2018.
- [2] Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum. Towards an iso standard for dialogue act annotation. In *Proceedings of LREC*, pp. 2548–2555, 2010.
- [3] Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, and David Traum. Iso 24617-2: A semantically-based standard for dialogue annotation. In *Proceedings of LREC*, pp. 430–437, 2012.
- [4] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. Dailydialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of IJCNLP*, pp. 986–995, 2017.
- [5] Chen Qu, Liu Yang, W. Bruce Croft, Johanne R. Trippas, Yongfeng Zhang, and Minghui Qiu. Analyzing and characterizing user intent in information-seeking conversations. In *Proceedings of SIGIR*, pp. 989–992, 2018.
- [6] Jan Ulrich, Gabriel Murray, and Giuseppe Carenini. A publicly available annotated corpus for supervised email summarization. In *AAAI08 EMAIL Workshop*, pp. 77–82, 2008.
- [7] William W. Cohen, Vitor R. Carvalho, and Tom M. Mitchell. Learning to classify email into “speech acts”. In *Proceedings of EMNLP*, 2004.
- [8] Amy X. Zhang, Bryan Culbertson, and Praveen Paritosh. Characterizing online communities using coarse discourse structures. In *Proceedings of ICWSM*, pp. 357–366, 2017.