

言外の情報に基づく雑談対話システム

2021年 3月

光田 航

言外の情報に基づく雑談対話システム

光田 航

システム情報工学研究群

筑波大学

2021年 3月

概 要

雑談対話は人間の会話の 60%を占めると言われており、相手のことを理解したり、悩みごとを相談したりなど、人間社会において重要かつさまざまな役割を担っている。コンピュータに人間との雑談対話を可能にさせる雑談対話システムの研究は、スマートスピーカーに代表されるロボットが今後人間社会に浸透し、よりさまざまなシーンで活躍するために重要である。このような理由から、雑談対話システムの研究が注目されている。

雑談対話システムが扱う対話の内容は多様であり、システムが人間のように協調的に振る舞うためには、ユーザの発言（ユーザ発話）の意図を的確に理解し、それに合った応答を返す必要がある。しかしながら、現在の雑談対話システムは、理解可能な対話の内容がユーザ発話に含まれる内容（例えば、キーワードなど）に限定されており、簡単な話題で話を続けることしかできない。人間は短いやり取りで協調的に振る舞うことができるが、これは発話に明示的に表れない重要な情報を推定し、多くのことを理解しているからに他ならない。今後、より高度な雑談対話、例えば意思決定など、人間と共同作業を行うような対話が可能な雑談対話システムを実現するためには、システムが発話に表れない情報も捉え、その情報に基づいて対話を行う技術が必要となる。

本研究では、ユーザ発話に明示的に表れない情報を考慮する雑談対話システムの実現を目指す。この、発話において明示されないが、対話において重要な情報を「言外の情報」と呼ぶことにする。言外の情報を考慮する対話システムが実現すべき対話の内容は明らかではないため、本研究では、アプローチを 3つのステップに分割する。まず、言外の情報の種類を明らかにするために、雑談対話における言外の情報の収集と類型化を行う。次に、言外の情報を雑談対話システムに応用することで、言外の情報の有用性を検証する。最後に、言外の情報を考慮する対話システムのタスク設定、および、対話システムの構築を行う。なお、本研究で扱う雑談対話のデータ、および、雑談対話システムは全てテキストベースであり、音声や画像など、テキスト以外の情報は扱わないことに注意されたい。

第 3 章では、雑談対話における言外の情報の収集、および、その類型化について述べる。これまで、会話の中で暗に伝わる情報について扱う研究が多数報告されているが、言外の情報を明らかにする試みは存在しない。本研究では、まず、人間同士の雑談対話のログを用いて、ログ中の各発話に対して、複数の作業者が言外の情報を記述することで言外の情報を大規模に収集した。その後、複数の作業者が相談しながら類似する言外の情報をまとめ上げていくことで、言外の情報の類型を作成した。作成された類型を評価するため、複数の作業者が言外の情報の事例に対してどの類型に該当するかをラベル付け（この作業をアノテーションと呼ぶ）した結果、作業者間のアノテーション結果が高い一致率（類型の単純一致

率で 0.75, Fleiss の κ 値で 0.69) となり, 作成された類型が妥当であることが確認できた。

第 4 章では, 雑談対話システムにおける言外の情報の有用性の検証について述べる。雑談対話システムにおいて言外の情報を明示的に扱うことを試みた研究は存在しないため, 言外の情報が雑談対話システムに有用であることを検証する必要がある。本研究では, 言外の情報を対話システムの発話理解結果とみなし, そのままシステム発話として発話した際に, ユーザに及ぼす効果について分析を行うことで言外の情報の有用性を検証した。雑談対話システムにおいて, システムの理解結果をユーザに伝えることでユーザに「システムに理解されている」という印象を持たせることは重要であるため, 言外の情報を用いて, この理解結果の伝達を実現できるかを検証した。第 3 章のデータ収集の過程で, 言外の情報の事例, および, その類型が得られているため, これらの知見に基づき, 類型ごとに言外の情報を伝達するシステム発話を作成, 評価した。その結果, ユーザの内面に言及する言外の情報はその伝達が好ましくない一方, ユーザと無関係な一般的事実に言及する言外の情報は極性によらずその伝達が好ましく, 雑談対話において有用ということが明らかになった。

第 5 章では, 第 3 章で述べた言外の情報の類型化, および, 第 4 章で述べた言外の情報の有用性検証から得られた知見に基づき, 言外の情報を用いる雑談対話の課題を設定し, 対話システムを構築した結果について述べる。言外の情報の類型化および有用性検証の結果から, 雑談対話では事実に関する言外の情報が有用であることが明らかになった。そのため, 言外の情報を考慮する雑談対話システムはまず, 事実に基づく対話であり, かつ, 今後人間との共同作業につながるような実用性を伴う対話を扱うことが適切だと考えられる。そこで本研究では, これらの知見を間接的に利用して, 第 3 章と第 4 章で扱った一般的な雑談とは異なる, オープンドメインの議論対話を設定した。議論対話は, 支持や反論の事実を言外の情報として述べることで合意形成を行う対話であり, 話者同士の協調的な振舞いで実現される代表的な共同作業であることから, 本研究で扱う対話として適切であると考えた。議論対話を実現するために, 入力されたユーザ発話に対して支持や反論を生成する発話生成手法を構築し, 5 点満点の人手評価において 3.93 点 (支持), および, 3.76 点 (反論) という高い正解率で議論の発話を生成できることを確認した。

本研究は, 雑談対話における言外の情報を明示的に扱うことを試みた最初の研究である。本論文の貢献は 3 点ある。1 点目は, 今後言外の情報を推論する対話システムを構築する際に有用と考えられる知見を示した点である。具体的には, 本論文が示す類型を用いることで, 言外の情報の推論を類型ごとに推論する部分問題に分けることができ, 既存技術 (例えば, ユーザプロフィールを推定する技術) の適用や, データセットの効率的な構築が可能になると考えている。また, 言外の情報の収集方法は本研究で扱った雑談対話以外にも適用できると考えられるため, 本研究の収集方法を用いて任意の対話で言外の情報を収集することが可能になる。

2点目は、言外の情報を雑談対話システムの発話理解結果とみなすことで、その伝達が好ましいもの、好ましくないものを明らかにする最初の試みを行い、その傾向を明らかにした点である。得られた知見は今後、雑談における発話理解結果を伝達する機能を有する対話システムを構築する際の有用な指針になると考えられる。3点目は、事実に関する言外の情報を扱う対話として、オープンドメインの議論対話タスクを提案し、支持や反論を伴う議論の発話を生成する手法を構築した点である。本手法は既存の対話システムのように、単に入力に関連する話題で応答するのではなく、言外の情報（支持や反論という論理的な関係）を扱うことができ、人間との共同作業のような高度な対話につながる技術であると考えている。

目次

第1章 序論	1
1.1 研究背景	1
1.2 先行研究の課題と本研究の位置付け	2
1.3 本研究の貢献	4
1.4 本論文の構成	9
第2章 関連研究	10
2.1 対話理論	10
2.1.1 言語行為論と対話行為分類	11
2.1.2 協調の原理とプラン認識	11
2.1.3 共通基盤	13
2.1.4 ポライトネス	13
2.2 対話システムの類型	15
2.2.1 達成すべき目標の有無	16
2.2.2 モダリティ	17
2.2.3 ドメイン	18
2.2.4 人数	18
2.3 雑談対話システム	18
2.3.1 ルールベースに基づく雑談対話システム	18
2.3.2 抽出ベースに基づく雑談対話システム	20
2.4 深層学習に基づく雑談対話システム	21
2.4.1 深層学習を用いた基本的な発話生成モデル	21
2.4.2 大量の対話データに基づく大規模発話生成モデル	23
2.5 言外の情報を扱う対話システムおよび言語処理技術	24
2.6 本章のまとめ	26
第3章 雑談対話における言外の情報の収集と類型化	28
3.1 言外の情報の収集	28
3.1.1 言外の情報の収集対象となる対話テキスト	28
3.1.2 言外の情報の収集手順	29
3.1.3 収集された言外の情報	30
3.2 言外の情報の類型化	32

3.2.1	言外の情報の類型化手順	32
3.2.2	言外の情報の類型	34
3.3	類型の評価	37
3.4	本章のまとめ	39
3.5	本章の付録	40
3.5.1	言外の情報を記述する作業への指示	40
3.5.2	言外の情報を類型化する作業への指示	43
3.5.3	対話理論に基づくトップダウン型の類型化の試み	44
3.5.4	言外の情報を類型化する過程で作成された全ての類型	47
第4章	雑談対話システムにおける言外の情報の有用性検証	48
4.1	有用性検証の流れ	48
4.2	評価対象となるシステム発話の種類	49
4.3	システム発話の作成と評価の方法	50
4.4	システム発話の主観評価結果	52
4.5	言外の情報の類型に基づく主観評価値の分析	53
4.6	本章のまとめ	58
第5章	言外の情報に基づく議論対話システムの構築	60
5.1	タスク設定	60
5.2	主張・論拠ペアデータの人手収集	61
5.2.1	収集手順	61
5.2.2	収集された主張・論拠ペアデータ	63
5.2.3	収集された主張・論拠ペアデータの分析	63
5.3	ガイドと主張・論拠ペアに基づく発話生成モデル	64
5.3.1	主張と論拠の立場を考慮するガイド	65
5.3.2	ガイド付き主張・論拠ペアの自動獲得	66
5.3.3	ガイド付き主張・論拠ペアの人手獲得	67
5.4	評価実験	68
5.4.1	評価対象の発話生成手法	68
5.4.2	モデルの学習と評価の方法	70
5.4.3	自動評価	71
5.4.4	人手評価	74
5.4.5	提案手法の誤り分析	77
5.5	本章のまとめ	78
5.6	本章の付録	79
5.6.1	主張・論拠ペアデータ収集のための指示と記述用フォーム	79
5.6.2	出力された論拠の評価に利用した指示	81

第 6 章	結論	82
6.1	本論文のまとめ	82
6.2	今後の課題	84
謝辞		87

第1章 序論

1.1 研究背景

雑談対話は人間の会話の約60%を占めるといわれており[114]，対話を通じてユーザが持つ特定の目標を達成するタスク指向型対話システムとは異なる，雑談対話システムの研究が注目されている[68, 27, 88, 1]．雑談対話システムの用途は非常に幅広く，タスク指向型対話システムの補助[113, 70]，ユーザがシステムに対して持つ信頼感の醸成[48, 31, 81]，カウンセリングを目的とした傾聴[49, 128, 34]，人間のような振舞いを再現したロボットの実現やそれを用いたコミュニケーション能力訓練[118, 108, 111]，魅力的なキャラクターの実現によるエンタメへの応用[99, 29, 106]など，さまざまなシーンへの応用が期待されている．

雑談対話の内容は多様であり，システムが人間のように協調的に振る舞うためには，ユーザの発言（ユーザ発話）の意図を的確に理解し，それに合った応答を返す必要がある．既存の雑談対話システムが理解可能な情報として，例えば，対話における話題を表す焦点語として用いられるキーワード[27]，ユーザの意図を表す情報として用いられる対話行為[35]，システムが出来事を理解するための情報として用いられる述語項構造[27]，ユーザの状態に適した行動を取るための情報として用いられる感情[48]，パーソナライズされた応答を返すために用いられるユーザ属性[32, 37]，発話の意味内容を表すベクトル表現[88, 1, 80]が報告されている．

既存の雑談対話システムにおける発話理解では，理解可能な内容が，発話に含まれている内容に限定されてしまっており，非常に単純な話題でしか対話を続けることができない．対話の理論研究における発話媒介行為（発話の結果間接的に引き起される行為であり，例えば説得など）[3]や，会話の含み（一般的な対話のルールから言語的に逸脱した方法で発話し，言外に意味を伝えることであり，例えば皮肉など）[22]で議論されているように，対話には，発話の内容に明示的に表れないが対話を進める上で重要な情報が存在する．人間は短いやり取りで協調的に振る舞うことができるが，これは発話に明示的に表れない重要な情報を推定し，多くのことを理解しているからに他ならない．今後，人間と共同作業を行えるような実用的な対話システムを実現するためには，システムが発話に表れない情報も捉えて対話を行う技術が必要となる．

そこで本研究では，ユーザ発話に明示的に表れない情報を考慮する雑談対話システムの実現を目指す．この，発話において明示されないが，対話において重要な情報を「**言外の情報**」と呼ぶことにする．言外の情報を考慮する対話システム

が実現すべき対話の内容は明らかではないため、本研究では、アプローチを下記の3つのステップに分割する。

- (1) 雑談対話における言外の情報の収集と類型化
- (2) 雑談対話システムにおける言外の情報の有用性検証
- (3) 言外の情報に基づく議論対話システムの構築

(1)として、言外の情報の種類を明らかにするために、雑談対話における言外の情報の収集と類型化を行う。これまで、会話の中で暗に伝わる情報について扱う研究が多数報告されているが、言外の情報を明らかにする試みは存在しない。(2)として、言外の情報を雑談対話システムに応用することで、言外の情報の有用性を検証する。言外の情報を対話システムの発話理解結果とみなし、そのままシステム発話として利用した際に、ユーザに及ぼす効果について分析を行う。雑談対話において、言外の情報を扱うことを試みた研究は存在しないため、言外の情報が協調的な対話に有用であることを検証する必要がある。(3)として、得られた知見に基づいて、対話内容の設定、および、システムの構築を行う。言外の情報の類型化、および、有用性検証の結果から、雑談対話では、事実に関する言外の情報が有用であったため、言外の情報を考慮する雑談対話システムはまず、事実に基づく対話であり、かつ、今後人間との共同作業につながる実用性を伴う対話を扱うことが適切だと考えられる。そこで本研究では、オープンドメインの議論対話を扱う。議論対話は、支持や反論の事実を言外の情報として述べることで合意形成を行う対話であり、話者同士の協調的な振舞いで実現される代表的な共同作業であることから、本研究の課題として適切である。

表1.1に、本研究のステップの概要とステップ間の関連を示す。本研究の内容は、大きく(1)から(2)まで、および、(3)で分かれる。これは、(1)と(2)が一般的な雑談対話を対象にした研究であるのに対し、(3)は(1)と(2)の結果に基づき、得られた知見を間接的に利用して雑談対話の内容を限定して対話システムの構築に取り組んだためである。

1.2 先行研究の課題と本研究の位置付け

本研究は大きく3つの内容に分かれるため、先行研究も大きく3つに分類することができる。本節では、それぞれの内容について代表的な関連研究を説明し、それらの課題、および、本研究の位置付けについて述べる。より詳細な関連研究については第2章で述べる。

表 1.1: 本研究のステップの概要とステップ間の関連

ステップ	対話の種別	検討内容の種別	前ステップとの関連
(1) 雑談対話における言外の情報の収集と類型化	雑談 (文脈あり)	データ収集, 分析	NA
(2) 雑談対話システムにおける言外の情報の有用性検証	雑談 (文脈あり)	システム発話の 作成, 分析	言外の情報の類型ごとにシステム発話を作成し評価
(3) 言外の情報に基づく議論対話システムの構築	議論 (文脈なし)	データ収集, 対話システム構築	有用な言外の情報の知見を間接的に利用して課題を設定

(1) 雑談対話における言外の情報の収集と類型化

雑談対話における言外の情報を類型化した試みは、我々が知る限りで存在しない。しかしながら、会話によって暗に伝わる情報についてはいくつもの試みがなされており、例えば、関連研究として、発話行為論 [3]、会話の含み [22]、プラン認識 [2] などが挙げられる。発話行為論 [3] における発語媒介行為は、その行為自体の意味が、解釈する話者に依存するため、結果として伝わる内容は言外の情報と考えることができる。また、協調的な会話からの逸脱から生まれる会話の含み [22] はまさに言外の情報の好例である。プラン認識 [2, 45] の研究では、協調的な対話を実現するために相手のプランを認識するが、これは相手から陽に伝えられていない意図や欲求を理解する必要がある、言外の情報を扱っていると考えることができる。これらの研究では、対話における言外の情報の伝達やその理解の過程がモデル化されている。しかしながら、その過程の結果、どのような情報が相手に伝わるのかについては明らかにされていない。雑談対話では、タスク指向型対話と異なり、話されている内容も幅広い。現状の技術では、プラン認識を行うことも難しいと考えられる。そうした中で、本研究では雑談対話システムが言外の情報を扱う第一歩として、雑談対話においてどのような言外の情報が見られるかに着目して類型を構築する。

(2) 雑談対話システムにおける言外の情報の有用性検証

雑談対話システムにおいて、発話理解結果をユーザに伝達したときの効果を分析した研究は、我々の知る限り見られない。しかし、発話一般について、相手に及ぼす効果を推定する研究は見られる。特に、ポライトネス理論 [8] に基づくものはいくつかある。ポライトネス理論とは、対話における発話の好ましさを、参加者の欲求（フェイス）という概念を用いて説明する理論である。フェイスにはポジ

ティブフェイス（聞き手に承認されたいという欲求）とネガティブフェイス（自由に行動を選択したいという欲求）の2種類があり、その両方に配慮することで対話が円滑に進むと考える．具体的な研究として、例えば、Wang らは、タスク指向型対話を対象に、発話が相手のフェイスを侵害するかどうかを判定する手法を提案している [94]．Walker らは、特定の状況の対話において、話者間の関係や対話行為に基づいて発話が相手のフェイスを侵害する度合いを推定し、発話を制御する手法を提案している [90]．Gupta らは、料理ドメインのタスク指向型対話システムに Walker らの手法を適用し、フェイスを侵害しない表現を生成する手法を提案している [24]．これらの研究に対し、本研究は、言外の情報に着目し、ユーザへの伝達において、ポライトネスを含むどのような要素が発話の好ましさに影響するかを総合的に分析するものであると位置付けられる．

(3) 言外の情報に基づく議論対話システムの構築

議論を扱う研究として近年最も代表的なものは、Argumentation mining[44, 79] というタスクである．これは、小論文や Web 上のフォーラムなど、意見や議論を含むテキストから、主張となる命題やその前提条件となる命題を抽出し、命題間の関係（例えば、支持や反論）を推定するタスクである．対話システムの分野では、ユーザと議論を行う対話システム [66, 30, 63] が提案されているものの、これらのシステムは人手で作成した議論の知識（例えば、意見をネットワーク状に繋いだ構造化された知識）を検索することで応答を返すため、さまざまなトピックへの対応が難しい．また、議論に関する発話生成の研究として、システムに入力された主張に対し、関連する議論の内容を生成する手法が提案されている [33, 59, 26]．しかし、これらの技術は対話システムへの適用を想定していないため、生成される議論の支持や反論を制御する機構が設けられておらず、対話システムの意図に応じて支持や反論を使い分けることができない．本研究は、さまざまなトピックを含むユーザ発話に対して、適切に支持や反論を生成するものであると位置付けられる．

1.3 本研究の貢献

(1) 雑談対話における言外の情報の収集と類型化

対話において人間が発話から理解する情報のうち、発話の内容とは異なる、明示されていなくとも人間が理解する情報、これを「言外の情報」(perceived information) と定義し、そのデータ収集と類型化を行った．

図 1.1 に、雑談対話における言外の情報を類型化する流れを示す．まず、言外の情報のデータ収集を行った．収集手順として、まず、言外の情報を記述する対象

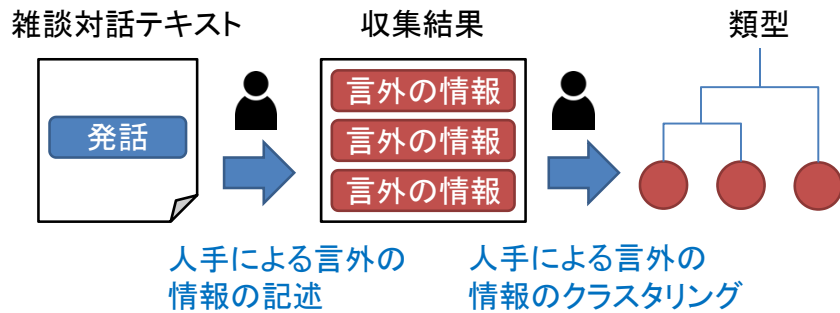


図 1.1: (1) 雑談対話における言外の情報の収集と類型化の流れ

となる対話テキストを用意した．次に，複数の作業者が対話テキスト中の各発話に対して言外の情報を記述することで，言外の情報を収集した．

対話テキスト中の各発話に対して，複数の作業者に言外の情報を記述してもらうことで，言外の情報を収集した．一般的かつ多様な言外の情報を網羅的に収集するために，作業対象として大規模な雑談対話コーパスを利用し，調査を行うのに十分な規模と考えられる，43 名の話者による 1,257 発話を含む 30 対話分の雑談対話テキストを利用した．これらのテキストに対し，6 名の作業者が独立に全テキスト中の発話に対して言外の情報を記述することで，計 11,533 事例の言外の情報を収集した．

収集された言外の情報を用いて，複数の作業者が人手で類似する事例をまとめ上げていくことで，言外の情報を類型化した．類型を作成する際には，特定の作業業者による恣意的な類型ではなく，一般性を持った類型が作成されるよう，複数の作業業者が類型化を行い，それらを統合するようにした．複数の異なる類型を用いて，その共通部分をまとめ上げることで，全ての作業業者の意見が反映された，一般性の高い類型が作成できると考えたためである．実際の作業は 1 グループ 3 名からなる計 6 グループが行った．

まとめ上げ作業の結果，レベル 1 からレベル 3 まで，階層的に言外の情報の類型が列挙された．最も細かい類型は 9 種類（願望 1，願望 2，信念 1，信念 2，属性，環境，行動，事実 1，事実 2）となった．作成された類型には，対話システムの内面のモデルとしてよく検討される Belief-Desire-Intention モデル [7] で扱われる信念や欲求に関する類型や，ユーザモデルの研究でよく検討される，ユーザ属性などのプロフィール情報 [31] が含まれていた．これら以外の情報として，話者の行動，環境，事実に情報といったことも重要であるという興味深い結果が得られた．作成された類型の妥当性を評価するために，言外の情報に対して人手でアノテーションを行い，作業業者間一致率を調査した．その結果，類型の単純一致率で 0.75，カッパ値で 0.69 と高い一致率でアノテーションが可能な類型であることを確認した．

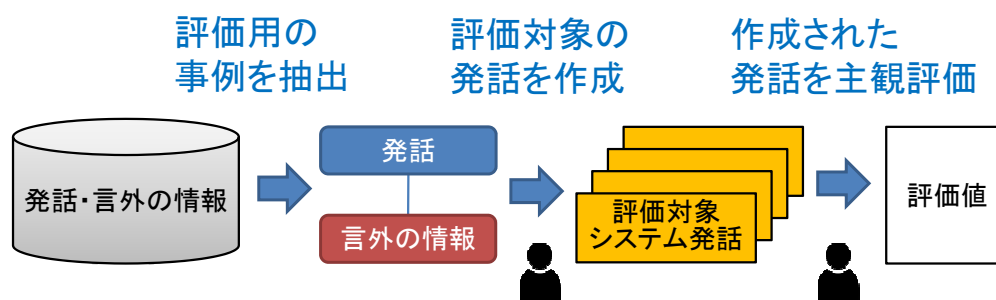


図 1.2: (2) 雑談対話システムにおける言外の情報の有用性検証の流れ

(2) 雑談対話システムにおける言外の情報の有用性検証

雑談対話システムにおける言外の情報の有用性を検証するために、最も基本的な方法として、理解した言外の情報をそのままシステムの発話として応答したとき、ユーザにどのような影響を及ぼすのか、また、雑談対話において有用な言外の情報が存在するのかを検証した。具体的には、雑談対話システムが発話から理解しうるさまざまな種類の情報をユーザに伝達した際の影響について分析を行い、自然さを損ねることなくシステムが発話可能な言外の情報を明らかにした。例えば、「気が向いたら、富士山に行きます」という発話からは、「ユーザは富士山が好き」ということや、「富士山は気軽に行ける」ということがわかるため、これを発話として伝達したときの影響を調査し、発話できる情報かどうか（この例の場合は発話可能）を検証した。さらに、ポライトネスに関する対話理論を用いて評価結果の説明を試みた。

図 1.2 に、検証の手順を示す。言外の情報を伝達するシステム発話を人手で作成、評価することで、異なる種類の言外の情報を伝達した際のシステム発話の自然さを被験者実験により調査した。言外の情報の種類として、最も細かい言外の情報の種類（レベル3）を利用した。比較対象として、雑談対話システムによる応答（自動応答発話）、ユーザ発話を繰り返す応答（繰り返し発話）、人手で作成した適切な応答（人手応答発話）を用意した。さらに、言外の情報の種類に応じて、システム発話がユーザに及ぼす影響の違いについて分析することで、どのような情報であればユーザに自然に伝達することができるのかを調査した。

評価の結果、言外の情報の伝達発話は全体としては繰り返し発話や人手応答発話と比べて自然さが低く評価されたものの、個別の種類としては大きく三群（低評価群、中評価群、高評価群）に分けられることがわかり、これら3つの群の間には、有意な差が見られた。

低評価群の発話は、願望2、信念2、願望1の発話理解結果を含んでおり、主にユーザの一時的な状態として、思考内容や感情に言及する発話である（例えば、「私に同意したいのですね」）。低評価群に含まれる発話は、対話中のユーザの一時的な状態に言及しており、ユーザの状態を一方的に決め付けるような発話と言え、ユーザのネガティブフェイス（自由に行動したいという欲求）を侵害していると

考えられる。

中評価群の発話は、属性、信念1、環境、行動の発話理解結果を含んでおり、主にユーザの興味、経験、プロフィールに言及する発話である（例えば、「カッコいい車が好きなんですね」）。この種類の発話は繰り返し発話と近い評価値が付与されていた。これらの発話はポジティブな内容を含んでおり、ユーザのポジティブフェイス（他者に承認されたいという欲求）を満たしているため高い評価値が付与されていると考えられる。

高評価群の発話は、事実1と事実2の発話理解結果を含んでおり、ユーザに直接関係のない一般的な事実に言及する発話である（例えば、「スポーツバイクは値段が高いですね」）。この結果は、一般的な事実に関する発話理解結果であれば、それを伝達するシステム発話は、自然さの評価値が高く人手で最適な応答を作成した場合と近いことを示す。

以上の被験者実験の結果をまとめると、対話中のユーザの一時的な状態（例えば、対話中のユーザの考え）はシステム発話として自然さの評価値が低く、一方でユーザ属性やユーザ自身と直接関係のない事実に関する情報はシステム発話として自然さを損ねることなく伝達することができることが示唆される。この結果から、言外の情報に基づく対話システムを構築するにあたって、まず事実に関する言外の情報を扱う対話に取り組むことが適切であると考えられる。

(3) 言外の情報に基づく議論対話システムの構築

雑談対話システムにおいては、言外の情報の種類のうち、事実1および事実2の種類の伝達発話の自然さにおける評価値が最大となり、これらの種類の言外の情報の伝達発話が有用なことが明らかになった。そのため、言外の情報を考慮する雑談対話システムは、事実に基づく対話、かつ、今後人間との共同作業につながる実用性を伴う対話を扱うことが適切だと考え、オープンドメインの議論対話を扱った。

本研究では、深層学習ベースの発話生成モデルと呼ばれるモデル（入力と出力をベクトルとして扱い、入力のベクトルを出力のベクトルに変換することで発話を生成するモデル）を議論の発話生成に利用する。発話生成モデルを作成するためには、所望の入力と出力をペアにした入力・出力ペアデータ（本研究の場合は主張・論拠ペアデータ）を大量に含むデータを用意し、入力から出力への変換をモデル内部のパラメータとして学習させる必要がある。一般にこのデータを学習データと呼び、学習データの品質は発話生成モデルの性能に大きく影響を与えるため、本研究の場合、様々なトピックについて、適切な支持や反論のペアが多数含まれていることが望ましい。

図1.3に、言外の情報に基づく議論対話システムの構築の流れを示す。本研究では、支持または反論のペアを正確に発話生成モデルに学習させるために、ガイドを導入した。ガイド付き主張・論拠ペアデータは、二種類の方法（自動収集と人手

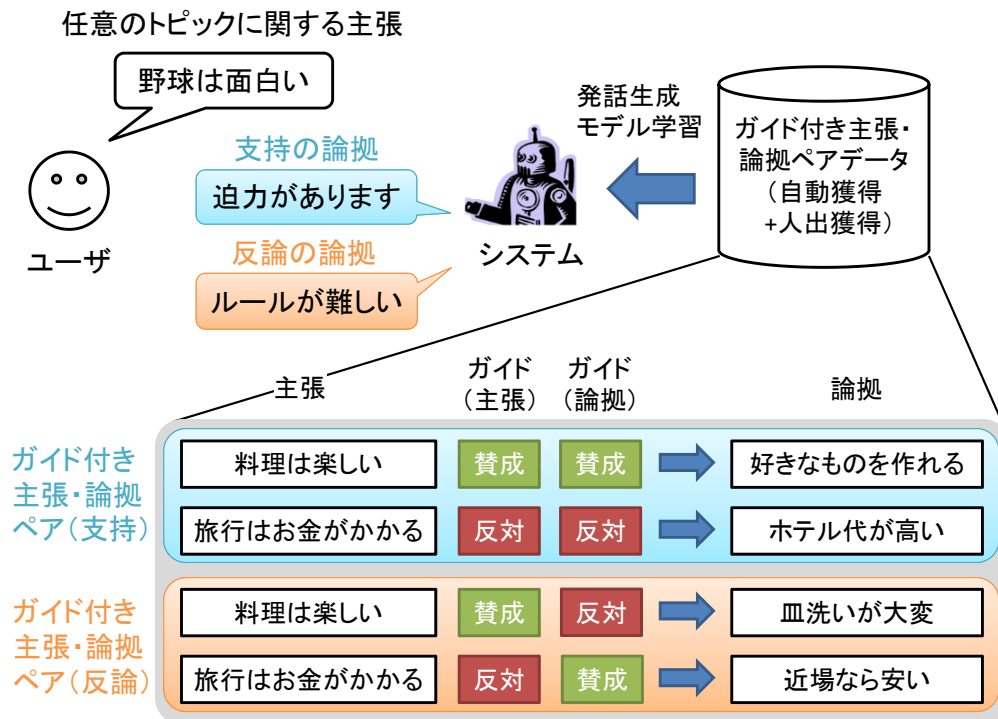


図 1.3: (3) 言外の情報に基づく議論対話システムの構築の流れ

収集)で作成した。自動収集では、大規模テキストから自動的に主張・論拠ペアデータを獲得する。品質は完全ではないものの、さまざまなトピックに関するペアを獲得することができた。もう一方の人手収集では、人手で主張・論拠ペアデータを書き下す。規模は限られるものの、高品質なペアを収集することができた。

本研究でガイドを導入したのは、入力となる主張がトピックに賛成する立場の発話(例。「料理は楽しい」)だけでなく、トピックに反対する立場の発話(例。「旅行はお金がかかる」)も含むためである。出力される論拠は、図に示すように、これらの立場を考慮し、支持であれば主張と同じ立場、反論であれば主張と異なる立場でなければならない。本研究では、この立場の情報をガイドとするトークン(発話生成モデルが扱う入力および出力に含まれるシンボルの最小単位)を用いて入力に追加することで、正確に支持や反論を生成する。

主張・論拠ペアの自動収集では、支持または反論の関係に近いと考えられる文のペアを大規模に獲得する。ペアを獲得する一般的な手法として、2文の関係を分類器で推定し、支持や反論の関係にある文を獲得する手法や、支持または反論を表す接続詞(「したがって」や「しかし」)で繋がる2文を獲得する手法が考えられるが、精度やカバレッジが十分とは限らない。そこで、本研究では、特定のトピックに言及する文であり、かつ、極性を持つ文をランダムにペアにすることで、尤もらしいペアデータを大規模に作成した。

主張・論拠ペアの人手収集では、クラウドソーシングを用いて、さまざまなトピックに関する主張、および、それに対する支持の論拠と反論の論拠を高品質に

なるよう収集した。各作業者は、＜トピック、主張、支持の論拠、反論の論拠＞の4つ組を3つずつ作成するようにした。データの収集後、自動収集手法と同様に、主張と論拠に対して評価表現抽出を行うことでガイドを追加し、ガイド付き主張・論拠ペアデータを作成した。

提案手法を人手評価した結果、5点満点の評価において、支持で3.93点、反論で3.76点となり、支持と反論の両方で評価値の有意な向上が確認できた。ベースラインと提案手法の間には、有意な差が見られた。この結果から、提案手法が論拠の生成に有効であることが確認できた。以上の結果から、言外の情報に基づく対話として、支持や反論の論拠を述べる議論対話において、本手法が有効であることが確認できた。本手法は部分的ではあるものの、言外の情報を扱うことができ、人間との共同作業につながる実用性があるものだと考えている。

1.4 本論文の構成

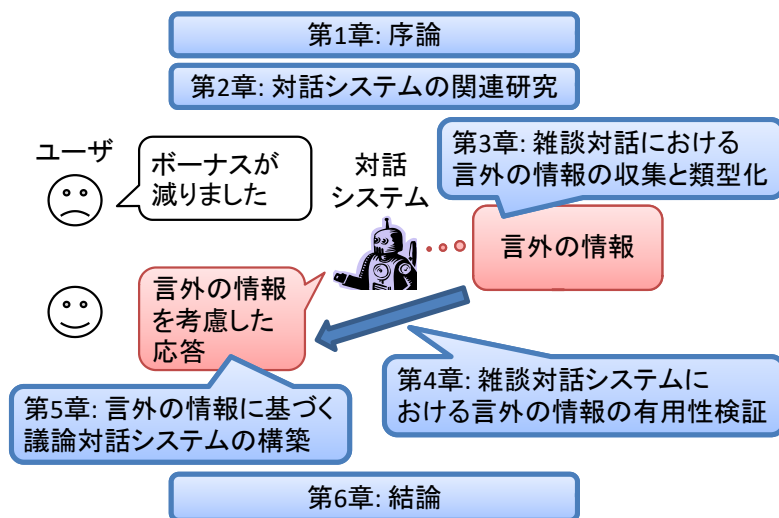


図 1.4: 本論文の章構成

図 1.4 に、本論文の構成を示す。第 2 章では、これまでの対話システムに関する研究、特に、雑談対話システムに関する研究を概観する。第 3 章では、雑談対話における言外の情報の収集手順、および、その類型化について述べる。第 4 章では、言外の情報の有用性を検証するために、言外の情報を対話システムの発話理解結果とみなし、そのまま発話に利用した際にユーザに及ぼす効果について分析を行った結果について報告する。第 5 章では、事実に関する言外の情報がシステムの応答として有効であるという知見に基づいて、事実に基づく雑談であり、かつ、今後人間との共同作業の実現につながる対話として、オープンドメインの議論対話システムを構築する手法について述べる。最後に、第 6 章にて、本論文の結論を述べる。

第2章 関連研究

本章では、対話システムの関連研究について説明する。まず、2.1節で対話の理論研究について述べる。次に、2.2節で対話システムの類型について説明した後、2.3節で、雑談対話システム（非タスク指向型対話システム）について説明する。その後、発展的研究として、2.4節で近年有効性が報告されている深層学習に基づく雑談対話システムについて述べる。最後に、2.5節で言外の情報に類する情報を扱った対話システム、および、言語処理技術について説明し、2.6節で関連研究のまとめを述べる。

2.1 対話理論

本節では、対話システムを構築する際の基礎となる対話の理論研究について説明する。人間同士で対話をする状況において、我々はあまり深く考えずとも、自然な対話をすることができる。例えば、自分が家でテレビを見ている際に、同居人に「お風呂が湧いたよ」と言われたとする。このとき、我々は意識せずとも非常に多くの情報を理解し、状況に応じた適切な応答を返すことができる。このような対話をシステムで再現しようとする場合、相手の発言からどのような情報を理解し、状況に合わせてどのような応答をすべきかを問題として定義する必要がある。このとき、対話の中で起きている現象を考慮することで、実際の対話に即した問題を設定することができる。この、対話の中でどのような現象が起きているかを理論的に明らかにすることを目的とするものが対話の理論研究であり、対話システム研究の基礎となる重要な理論である。

対話研究において、自然言語で発せられた情報のまとまりを**発話**（utterance）と呼ぶ。また、複数の発話から構成されるまとまりにおいて、独話を含めたものを**談話**（discourse）と呼び、二人以上の話者によるものを**対話**（dialogue）と呼ぶ[119]。独話や対話はしばしば**文脈**（context）とも呼ばれ、また、対話は**会話**（conversation）とも呼ばれる[121]。本論文は対話システムを対象とする研究であるため、対話システム分野で慣習的に用いられる対話という用語を以降用いる。

2.1.1 言語行為論と対話行為分類

対話の理論研究における最も基本的な理論として、発話とは何であるかを議論したものが Austin, および, Searle の**言語行為論**である [3, 74]. 言語行為論において、発話を行うことは特定の行為を行うことであると議論されている. 例えば、市長が「ここに市民体育大会を開催します」と発話した場合、これは市民体育大会が開催されたことを記述しているのではなく、発話自体が市民体育大会の開会という行為を遂行しているというものである [119]. Austin は、この発話から引き起こされる行為を**発話行為** (locutionary act), **発語内行為** (illocutionary act), **発語媒介行為** (perlocutionary act) の3種類に分類している. 発語行為は、特定の情報を伝えようとして、それを発話として話したり記述したりする行為そのものを指す. 発語内行為は、先の市長の例のように、発話によって直接的に引き起こされる行為（先の例の場合は開会という行為）を指す. 発語媒介行為は、発話によって間接的に引き起こされる行為全般を指し、例えばある発話を通じて相手を説得したり、笑わせたり、泣かせたりする行為を指す.

言語行為論における発話行為の分類体系は、実際の対話の分析にはそのままでは不十分であることが知られており [121], 対話の目的に応じて、分析に適切と考えられる類型を議論した研究が多く存在する [35, 9, 50, 98]. これらの類型は発話行為とは異なる用語で**対話行為** (dialogue act) の類型と呼ばれる. 代表的なものとして、SWBD-DAMSL (Switchboard Discourse Annotation and Markup System of Labeling) [35] は、Switchboard コーパス [21] (電話を用いた音声対話を収録したコーパス) を対象に設計された対話行為の類型であり、各発話を 42 種類の類型に分類している. また、ISO の標準規格 (ISO 24617-2) として対話行為のアノテーション規格 [9] が承認されている. ISO 24617-2 では、汎用の目的で利用可能なアノテーションの枠組みが定義されており、発話に対して付与される 56 種類の対話行為 (汎用で付与可能なものと、特定の目的を持つ発話に限定して付与可能なものに分かれる) が定義されている. これらの類型は汎用性に富むものの、対話システムでの利用を考えた際には粒度が細かいため、実際にはこれらの類型をベースに目的に合わせた類型を設計、利用することが多い. 例えば、Williams ら [98] は、レストラン案内など特定のタスクを達成する対話システムを対象に、Request や Inform などの対話行為を、伝達される情報 (レストランのジャンルや場所など) の定義とともに類型化している. また、Meguro ら [126] は、本論文と同じ日本語の雑談対話を扱っており、雑談対話中の発話の意図を、挨拶、情報提供、自己開示、質問などに分類している.

2.1.2 協調の原理とプラン認識

言語行為論では、対話を構成する発話を対象に、その役割について議論が述べられてきた. その後の発展として、発話単体や、発話のまとまりである対話にお

いて、話し手から聞き手に対して情報が伝達される過程を明らかにしようとする理論が構築されてきた。

対話における発話を通じて、発話の中に含まれる言語的な情報や情報、さらにその言語的な内容を越えた情報が伝わる枠組みを議論したのが Grice である [22]。例えば、ドライブ中に同乗者が「あそこにガソリンスタンドがあるよ」と発話したとき、聞き手はそのガソリンスタンドがおそらく営業しているという情報を理解することができる。このように、発話で明示されていなくとも聞き手が推論することで理解する情報を説明する理論として、Grice は**協調の原理**を提案した。協調の原理とは、会話の中で発言するときには、それがどの段階で行われるものであるかを踏まえ、また自分の携わっている言葉のやりとりにおいて受けいれられている目的あるいは方向性をふまえた上で、当を得た発言を行うようにすべきであるというものである。Grice はこの原理を守るために必要な行為の基準として、4 つの具体的な**格率** (maxim, しばしば公準とも呼ばれる) を定義している。その四つとは、量の格率 (要求に合わせて、必要かつ十分な情報のみを発言すること)、質の格率 (十分な証拠がない情報や、嘘の情報を発言しないこと)、関係の格率 (関連性のある発言をすること)、様態の格率 (曖昧な表現や不明瞭な発言は避け、わかりやすい発言をすること) である。Grice は、対話においてこれらの格率を守ることによって、発話の内容が伝わるとしているが、一方でこれらの格率を守らないことで、発話の内容を越えた情報が伝わると述べている。例えば、「X 嬢はホームスイートホームの楽譜に近い音声を発した」[22, 112] という発話は、「X 嬢はホームスイートホームを歌った」というわかりやすい表現の発話ではないため、様態の格率を違反している。しかし、格率を違反しているからといって情報が伝わらないのではなく、聞き手は、話し手があえて様態の格率を違反する言い方をしているということから、言語的な情報を越えた情報 (**会話の含み**) があることを慣習的に推論し、X 嬢の歌が上手ではなかったということを理解する。

対話において、協調の原理を守った対話がなされることで、我々は少ないやり取りからでもさまざまな情報を理解することができる。これは、協調の原理に従って、話し手の要求に応じた関連のある情報が対話を通じて意図として聞き手に伝わるためであり、対話と意図の関係を扱うプラン認識 [45] という理論で議論されている。プランとは、特定の目的を達成するために必要となる行為の系列を表すものであり、例えば電車に乗って目的地に辿り付くために必要な行為の系列や、ハンバーガー屋に言ってハンバーガーを食べるために必要な行為の系列などである。我々が特定の目的を達成するために必要な行為の系列を考えることを**プランニング**と呼び、また、対話を通じて話し手の意図、すなわち、話し手が今後実行しようとしているプランを聞き手が推論して理解することを、**プラン認識** (plan recognition) と呼ぶ。行為の系列を表すプランはそのままでは行為の系列であるため、対話を通じて、相手にプランを認識させ、必要となる行為 (例えば、食べたいハンバーガーを店員に伝えるという行為や、店内で食べるか持ち帰るのかを伝えるという行為) を順に達成していく必要があり、対話と深く関係している。このプランと対応す

る対話を**談話プラン** (discourse plan) と呼び、Litman と Allen によって、プラン認識と談話プランの関係が整理されている。

2.1.3 共通基盤

先に述べた対話の理論研究から、発話や対話を通じて、それらの中で言語的に明示された情報を越えて、さまざまな情報が慣習や推論の知識から理解されることが明らかになった。このような情報の伝達は、対話の参加者の間で慣習や推論の知識が共有されているために可能になっており、この共有されている知識を**共通基盤** (common ground)、または、**相互信念** (mutual belief) と呼ぶ [12]。共通基盤が正しく構築されていない場合、参加者の間で対話がかみ合わず、正確に情報をやり取りをすることが難しくなる。例えば、コメディドラマなどでよくある笑い話として、ある人物 A が友達 B に恋人 C を紹介する状況において、友達 B が誤って恋人 C を人物 A の親だと勘違いしたとする。このような状況で、人物 A が恋人 C と親密な様子を友達 B に伝えたとなると、友達 B は人物 A が意図せずとも、非常に気味悪い思いをする。これは、人物 A と友達 B の間で、紹介された人物 C が恋人であるという情報が共通で認識されていないためである。なお、共通基盤が正しく構築されていない場合、必ずしも対話がかみ合わないわけではない。先のコメディドラマが成立するのは、偶然的に対話がかみ合うことで、友達 B の反応が人物 A の意図するものと異なることを視聴者が面白く感じるためである。対話を通じて共通基盤が形成されること、例えば、友達 B に人物 C が人物 A の恋人であるという情報が正しく伝わることを**基盤化** (grounding) と呼ぶ。対話の各参加者が対話の中で有しているさまざまな知識 (常識、慣習、対話から理解し内容など、あらゆる情報) は**信念** (belief) と呼ばれ、参加者が共通して持つ信念が共通基盤である。我々が対話を通じてうまくやり取りができるのは、共通基盤がうまく形成できなかったとしても、その齟齬を解決するためのやり取り (例えば、問い返しや承認) を行っているからであり、このような対話を適切に行うために行う対話を**付带的コミュニケーション** (または、メタコミュニケーションやメタ対話) と呼ぶ。

2.1.4 ポライトネス

言語行為論における発語媒介行為や、協調の原理におけるの会話の含みは、対話を通じて言外に情報が伝わる枠組みについて整理している。一方で、なぜそのような枠組みが存在するのかという理由を明らかにしているわけではない。この理由を間接的に伝えることによる丁寧さ、**ポライトネス**という概念で説明したものがポライトネス理論 [8] である [124]。

ポライトネス理論とは、円滑な人間を確立・維持するため [109] に必要となる言語的な行為を整理した理論である。ポライトネス理論は**フェイス**と呼ばれる、対

話の参加者が持つ欲求に基づいて議論される。フェイスにはポジティブフェイス（聞き手に承認されたいという欲求）とネガティブフェイス（自由に行動を選択したいという欲求）の2種類があり、その両方に配慮することで対話が円滑に進むとされる。これら2種類のフェイスを脅かすような行為はFace Threatening Actと呼ばれ、話し手と聞き手の社会的距離、話し手と聞き手の権力差、行為自体の押しつけがましさの三つに基づいてフェイスを脅かす度合いが算出される。ポライトネス理論では、算出されたFace Threatening Actの度合いに応じて、相手に行う言語的な行為を変えるための方法論が定義されており、脅かしの度合いが低い行為（例えば、挨拶）であればそのまま発話する一方で、度合いが高い行為（例えば、命令）であれば婉曲的に発話する必要があるとしている。

ポライトネス理論は、対話システム構築において、システムがユーザと円滑に対話を行うための知見として利用されており、これまでにさまざまな研究が報告されている。具体的な研究として、Walkerらは、特定の状況の対話において、話者間の関係や対話行為に基づいて発話が相手のフェイスを侵害する度合いを推定し、発話を制御する手法を提案している[90]。Guptaらは、協調的に料理を行う状況における対話システムにおける発話生成にWalkerらの手法を適用し、フェイスを侵害しない表現を生成する手法を提案している[24]。Wangらは、ユーザに数学など特定の科目に関する問題のチュータリングを行う対話システムを対象に、発話が相手のフェイスを侵害するかどうかを言語的な表現（例えば、謙遜表現、強調表現、侮蔑表現など）に基づき判定する手法を提案している[94]。Niuらは、近年注目されている深層学習を用いた発話生成において、生成される発話のポライトネスの度合いを制御する手法を提案している[55]。また、対話システムを対象とする研究ではないが、Danescuらは、インターネット上の書き込みに対してポライトネスの度合いを付与したコーパスを構築し、その度合いを推定する手法を提案している[16]。

ポライトネスに関する別の理論として、神尾の**情報のなわ張り理論**[115, 116]がある。BrownとLevinsonのポライトネス理論では、主に話し手と聞き手の関係性（社会的距離や権力差）に着目してポライトネスが説明されている一方で、情報のなわ張り理論では、対話の中で伝達される情報と対話の各参加者との関係性に着目してポライトネスを説明している。

表2.1に示すように、なわ張り理論では、対話を通じて伝達される情報を、その情報と話し手（または聞き手）との心理的距離によって近いもの、遠いものという二つの集合に分類し、心理的距離の近いものをなわ張りの内、心理的距離の遠いものをなわ張りの外と定義している。なわ張りは話し手と聞き手のそれぞれに対して定義されるものであるため、対話において伝達される情報は、話し手のなわ張りの内か外か、また、聞き手のなわ張りの内か外かで4通りに分類される。なわ張り理論では、これらの4通りの種類の情報が話し手によって判断されており、日本語における文末表現として表れるとしている。表の例では、話し手のなわ張りの内は直接形（断定的な表現）、なわ張りの外には間接形（断定を避けた表現）、

表 2.1: 情報のなわ張り理論における 4 種類の情報の例と文末形式

話し手の なわ張り	聞き手の なわ張り	発話の例	文末形式
内	外	昨日はパンを食べた	直接形
内	内	今日はいい天気だね	直接ね形
外	内	最近君は元気そうだね	間接形ね形
外	外	昔の日本人は一日 2 食だったらしい	間接形

聞き手のなわ張りの内には助詞のね（同意または確認を表すもの）が用いられている。このとき、話し手のなわ張りの内の情報は直接形で表現することができるが、なわ張りの外の情報は間接形で表現しないと、聞き手に好ましくない印象（押しつけがましさ）を与えている。話し手のなわ張りの内に入る情報として、神尾は以下の 8 種類の情報を定義している [115, 110].

- (1) 話し手自身が直接体験によって得た情報
- (2) 話し手自身の過去の生活史や所有物についての個人的事実を表す情報
- (3) 話し手自身の確定している行動予定および計画などについての情報
- (4) 話し手自身の近親者またはごく身近な人物についての重要な個人的事実を表す情報
- (5) 話し手自身の近親者またはごく身近な人物の確定している重要な行動予定、計画などについての情報
- (6) 話し手自身の職業的あるいは専門領域における基本的情報
- (7) 話し手自身が深い地理的関係を持つ場所についての情報
- (8) その他、話し手自身に何らかの深い係りを持つ情報

2.2 対話システムの類型

本節では、対話システムの中の代表的な類型について説明する。対話システムに該当する技術の範囲は幅広く、音声、画像、言語など複数の分野にまたがる技術を指す場合も存在する。以降、代表的な類型 [121] として、達成すべき目標の有無、モダリティ、ドメイン、人数について述べる。

2.2.1 達成すべき目標の有無

達成すべき目標の有無とは、ユーザが対話を通じて達成したい明確な目標を持つか否かを指し、目標を持つ場合を**タスク指向型対話システム** (task-oriented dialogue system), 目標を持たない場合を**非タスク指向型対話システム** (雑談対話システム, chat-oriented dialogue system, chatbot) と呼ぶ。これらの類型は対話システムの設計を考える際に重要であり、タスク指向型対話システムでは、一般にその目標をなるべく頑健に達成できるようにシステムを設計するのに対し、雑談対話システムでは、特定の目標を仮定することが難しいため、一般になるべく自然にユーザと対話を継続させたり、ユーザの満足度を高めるようシステムを設計する。既存の対話システムの技術では、これら二つの類型の対話システムの両方の能力を十分に兼ねそなえたもの（例えば、SF 映画に出てくるような、自然な雑談をしつつ、必要に応じて雑談の内容を踏まえて目標を達成できるコンシェルジュのようなシステム）の実現は難しい。そのため、既存のシステムは二つのタイプのどちらかに分類できるか、タスク指向型対話システムに雑談対話システムの機能を付加したものと言える。

タスク指向型対話システムは、究極的には、目標が達成できればよいという点で、対話は最小限でよいと考えられる。また、ユーザが対話を開始した時点で特定の目標を意識していない場合も存在し、システム側が対話を通じて目標を明確化させてその目標を達成するもの（例えば、ユーザが検索したい内容に対話的に絞り込んでいく対話型検索エンジン）も該当する。タスク指向型対話システムは、目標が達成できればよいという位置付けから、検索や案内を行うシステムのインタフェースとして使用されることが多く、他のインタフェース（例えば、コンピュータ上でのボタンやリストボックスなどのグラフィカルユーザインタフェース）と競合してきた [120]。タスク指向型対話システムをインタフェースとして使用するときのメリットとして、ユーザが複雑なインタフェースの知識がなくともシステムを使えたり、対話を通じて目標を明確化できたりという点が挙げられる。しかし、競合のインタフェースの代替として利用するためには高い精度が求められるため、これらのメリットの実現には技術的な課題も多い [36]。

雑談対話システムは、特定の目標が存在しないため、対話すること自体が目的となる。そのため、情報案内や検索などの明確な目標タスクを設定することが難しい場合（例えば、議論を行う対話システム）や、タスクが達成できたかどうか外部から容易に判断できない場合（例えば、カウンリングを行う対話システム）に、雑談対話システムに分類されることも多い。タスク指向型対話システムと比較して明確な目標がないものの、目標が全くないということではなく、一般に抽象的なさまざまな目標（例えば、信頼感の醸成、情報の共有、思考の整理などの目標）を持つと考えられている。既存の雑談対話システムは、大きくユーザと自然に対話を続けることを目的とするシステムと、特定の状況の雑談にシステムを特化させたもの（例えば、傾聴を行うシステム）が存在する。いずれにせよ、ど

これらのシステムも明確な目標を仮定できず、システムを作り込むことが難しいため、高度な雑談対話システムの実現には多くの技術的課題が存在する [123]。技術的課題は多いものの、近年の対話システムの利用シーンの多様化の影響から、家庭内で利用するコミュニケーションロボット、魅力的なキャラクターとのやり取りを楽しむエンターテインメント、面接や恋愛などの対話に関する能力向上のための練習相手、対話を通じたカウンセリングなど、さまざまな分野への応用が期待されている。

2.2.2 モダリティ

対話システム研究において、**モダリティ**とは、システムが扱う入出力の情報の種別を指す。言語学や自然言語処理においては、モダリティは内容に対する話し手や書き手の印象（例えば、「雨が降るらしい」の「らしい」）を表すが、このモダリティとは異なる概念である。対話システムにおけるモダリティとして現在代表的なものは音声とテキストであり、音声は音声を通じた入力や出力を行う対話システム、テキストはテキストを通じた入力や出力を行なう対話システムを指す。音声を入出力に用いるシステムは一般に**音声対話システム**（Spoken dialogue system）と呼ばれる。テキストを入力として用いるシステムは**チャットボット**と呼ばれることが多く、このような呼び方をする場合雑談対話システムとして用いられることが多いが、タスク指向型対話システムを指す場合もある。一種類のモダリティを用いるシステムをシングルモーダルなシステムと呼ぶ。一方で、画像と音声のように、複数のモダリティを用いるシステムをマルチモーダルなシステムと呼ぶ。人間は対話を行なう際に、何が発話されたかという言語的な情報だけでなく、話者の声色、表情、身振りなどさまざまな情報を取得し、それらの情報に基づき対話を進めるため、マルチモーダルなシステムの実現は自然な対話を行なう上で重要な課題である。

モダリティごとの扱いの難しさとして、テキストのみ、音声のみ、音声と画像の順に高度な処理が要求される。テキストの場合は、一般にテキストチャットなどを利用してユーザが入力するため、タイポがある場合はあるものの、どちらが話すかというターンの決定や、入力された内容の正確な取得が容易である。音声の場合は、ターンの決定が難しく、言い間違いやノイズの影響による聞き間違い（音声認識エラー）が起きることも多く、声色などの重要な情報を含んではいるものの、テキストのみの場合と比較してより高度な処理が必用となる。さらに画像をモダリティとして含む場合も同様であり、画像からは表情や身振りなどさまざまな情報を取得できるものの、それらをテキストや音声と関連させてシステムに理解させるためには高度な処理が必用となる。

2.2.3 ドメイン

ドメインは、対話システムが扱うことができる話題を指す。例えば、レストラン検索ができる対話システムであればドメインはレストランとなり、電車の乗り換え案内ができるシステムであればドメインは電車となる。レストランドメインの対話は、電車の乗り換え案内の対話はドメイン外となり対応ができない。単一のドメインのみを扱うシステムを**シングルドメイン**と呼び、複数のドメインを扱うことができるシステムを**マルチドメイン**と呼ぶ。タスク指向型対話システムは一般に、目標を達成するために必用な対話が単一または数種類のドメインでカバーできる。このように、特定のドメインのみを扱うというシステムを**クローズドドメイン**と呼ぶ。一方、ドメインを限定しないものを**オープンドメイン**と呼び、こちらはさまざまな話題を扱う必用がある雑談対話システムに多い。ただし、特定の話題についてのみ雑談を行なうことができるシステムも存在し、そのようなシステムはクローズドドメインとなる。

2.2.4 人数

人数は、対話に参加する話し手または聞き手の人数を指す。一人の場合は独話となり、二人以上の場合是对話となる。既存の対話システムはシステムとユーザの二人を想定して作られているものが多い。システムとユーザを合わせて三人以上の対話はマルチパーティー対話と呼ばれる。二人の場合と比較して、次に誰が話すかといったターンの判定や、ユーザの発話が誰に向けての発話なのかといった聞き手の同定など、対話の内容はより複雑になる。

2.3 雑談対話システム

本章では、雑談対話システムの構築に関して代表的な研究を説明する。近年の雑談対話システムは深層学習を用いるものも多く、それらについては次章で説明し、本章ではそれ以外のルールベースシステム、および、抽出ベースシステムについて説明する。

2.3.1 ルールベースに基づく雑談対話システム

雑談対話システムは一般にタスク指向型対話システムと異なり、幅広い話題を扱う必用があるため、それらをカバーするためにルールベースの手法に基づいて研究が始まった。最も代表的なものはELIZA（イライザ）[96]であり、心理療法を行なうセラピストをシミュレーションした対話システムである。ELIZAは患者であるユーザの発話に対して、応答のテンプレート集から適切な応答を選択する

ことで対話を行なう。例えば、特定のキーワードを発見した場合はそのキーワードにあらかじめ紐付けられた応答を返し、特定のキーワードが発見できなければ、汎用的な応答を返すというものである。下記はその例である [96, 121].

ユーザ：男なんてみんな一緒
ELIZA：どんな風に？
ユーザ：いつもいろんなことでいらいらさせるの
ELIZA：具体的な例が思い浮かびますか？
ユーザ：うーん、私のボーイフレンドがここに来させたの
ELIZA：あなたの彼がここに来させた
ユーザ：彼は私がたいてい鬱状態だっていうの
ELIZA：あなたが鬱状態と聞いてお気の毒に思います

この例では、テンプレート集を用いて、特定のキーワードが発見できなければ汎用的な応答（質問など）を返し、特定のキーワード（鬱状態）が発見できれば、そのキーワードとテンプレートを用いた応答を返している。システムは対話の内容を理解しているわけではないものの、応答がかみ合えば自然な対話ができているようユーザは感じられる。

ELIZA をベースに、ルールの記述方法やそのインタプリタ（ルールを解釈して対話するシステム）を発展させたシステムが A.L.I.C.E（アリス）[91] であり、そのルールの記述方法を Artificial Linguistic Markup Language（AIML）と呼ぶ。近年の一般的な商用の雑談対話システムでは、カスタマイズ性の高いルールベースの手法が利用されることが多く、AIML がよく利用されている [127].

図 2.1 にルールベースの雑談対話システムが利用する対話ルールの例を示す。対話システムは、ユーザ発話を pattern にマッチさせることで発話を理解し、マッチしたルール（category）の template を返すことでユーザに応答する。雑談対話で扱いたい対話の内容に合わせてあらかじめ人手でルールを用意しておくことで、ルールの範囲内で品質の高い応答を実現することができる。AIML には pattern タグを含むさまざまなタグが定義されており、例えば pattern にマッチする発話の一つ前の発話（一般的にはシステム発話）にマッチする that タグが存在し、文脈を考慮したルールを記述することもできる。

幅広い話題で高品質な雑談を行うためには、大規模な AIML 構築が必要となる。日本語における大規模な AIML を作成し、雑談対話システムを構築した研究が Higashinaka らによって報告されている [27]. 対話コーパスを元に、連続する適切な 2 発話のペアを pattern と template のペアとみなすことで AIML の初期ルールセットを作成し、評価セット中の発話に対して適切な応答が返るまで AIML を修正することで、AIML の構築を行っている。また、250 個のキーワードに対して作成された 25,000 発話を元に、キーワードと発話のペアを pattern と template のペアとみなして AIML の拡充を行うことで、最終的に 149,300 個の category で構成される AIML を構築している。

```

<aiml>
  <category>
    <pattern>お酒 * 飲め * か</pattern>
    <template>飲めますよ</template>
  </category>
  <category>
    <pattern>タバコ * 吸 い * か</pattern>
    <template>全然吸いません</template>
  </category>
</aiml>

```

図 2.1: AIML の例. pattern は形態素単位で区切られており, ‘*’ が任意の形態素列と対応することでユーザ発話とマッチする.

ルールベースの雑談対話システム構築の課題として, 文脈まで含めたルールを記述することが難しい点や, ルールの規模が大きくなった際の調整が難しい点が挙げられる. 例えば, AIML の改善を行う試みとして, 不適切な発話の除去 [127] や, ルールの追加・修正 [69] を行った研究が報告されている. ルールの追加・修正では, 複数の作業者が作成したユーザ発話, および, 複数ターンの対話を通じてシステムの評価を行い, その評価値が閾値以上になるまで繰り返しルールの追加や修正を行うことで, AIML の改善を行っている. しかし, 大規模な AIML の改善は容易ではないことが報告されている. その要因として, 対応可能な話題が少ないためにシステムが同じ発話を繰り返してしまう点や, 文脈理解が必要な対話に対応できていない点が挙げられている.

2.3.2 抽出ベースに基づく雑談対話システム

図 2.2 に, 抽出ベース対話システムの概要を示す. 抽出ベースの手法は, あらかじめ入力発話と出力発話のペアを大量に獲得しておき, ユーザ発話に類似する入力発話を含むペアを検索してその出力発話を応答として返す手法である. 抽出ベースの手法はしばしば用例ベースとも呼ばれる. このペアを用いる代表的なシステムとして, Ritter らの Twitter ベースの対話システム [68] がある. Ritter らは, Twitter から獲得したリプライのペア集合を統計的機械翻訳のパラレルコーパスとみなし, 入力発話から出力発話への翻訳を行なうことで応答を返すシステムを提案している.

抽出ベース対話システムの構成要素は大きくリプライペアの構築, および, リプライペアの検索の二つに分けることができる. リプライペアを獲得する方法として, 映画の字幕を利用する手法 [5], テキスト解析の結果に基づく述語項構造 (主

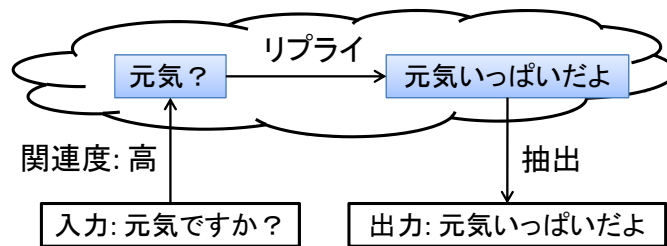


図 2.2: 抽出ベース対話システムの概要。図は文献 [122] より引用。

語や述語からなる文構造) を利用する手法 [100, 28], 特定のキャラクタのファンコミュニティが集合知的にリプライペアを作成する手法 [29] などが提案されている。一方, リプライペアを検索する手法としては, 一般的なキーワード検索である Lucene¹ を利用する手法や, 単語の異表記を吸収することができるベクトル検索 [51] を利用する手法が提案されている。抽出ベースの課題として, 出力があらかじめ決まったリプライペアの中から選択されるため, 文脈や入力発話に対して細かい表現の制御が難しいという点がある。

2.4 深層学習に基づく雑談対話システム

本節では, 近年雑談対話システム構築における主流の方法として盛んに研究されている深層学習に基づく雑談対話システムについて説明する。これまで述べてきたように, 雑談対話システムが扱わなければならない話題や表現は多様であるため, あらかじめ応答を用意しておくアプローチは難しい。この問題を解決する手法が深層学習に基づく手法であり, この手法では一般に, 大規模な対話データを用意し, 文脈を含む入力発話に対して対話データ中の出力発話を生成するよう発話生成モデルを学習させることで自然な対話を可能にする。大量の対話データを用意できれば, あらかじめシステムの応答を用意しておく必要がないため, 文脈や入力発話の内容を考慮した自然な応答が可能となる。

2.4.1 深層学習を用いた基本的な発話生成モデル

図 2.3 に, 発話生成モデルとして最も代表的な **sequence-to-sequence (seq2seq)** [89] の例を示す。この深層学習に基づく手法は機械翻訳の分野 [4, 83] で発展した後, 雑談対話システムへと応用された。図には基本的な一問一答の発話生成モデルが学習済みであったとき, 「趣味は何」という入力発話に対して「野球です」という出力発話を生成する例が示されている。発話生成モデルにおいて, 入力発話は**形態素**(単語として意味を持つ最小の単位のこと) または**サブワード** [75] (シス

¹<https://lucene.apache.org/>

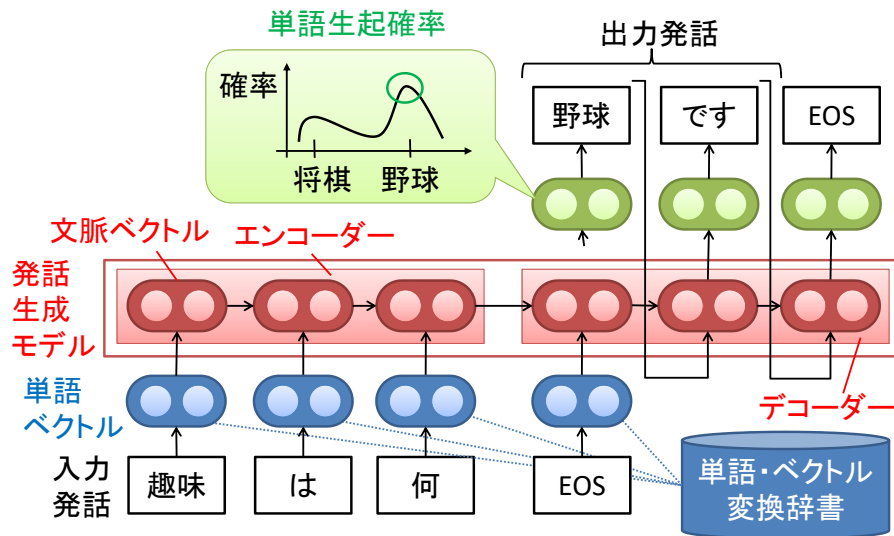


図 2.3: 深層学習に基づく雑談対話システムの例

テムが利用する辞書に登録されている単語または文字になるまで、発話を分割したもの)という単位で扱われる。形態素とサブワードは異なるものであるが、深層学習に基づく雑談対話システムのアルゴリズムを説明する上では共に**トークン**(または単語)として扱うことができるため、以降トークンまたは単語と呼ぶ。

学習済みの発話生成モデルは、まず、単語に分割された入力発話から、各単語と一対一で対応するベクトルを作成する。この単語からベクトルへの対応付けには単語・ベクトル変換辞書が利用され、この辞書は発話生成モデルと同時に学習されることが多いが、あらかじめ大規模なコーパスに基づいて学習された Glove[60]などのベクトル辞書を用いる場合もある。ベクトルに変換された入力発話中の各単語は、発話生成モデルのエンコーダと呼ばれる系列変換モデルに入力される。この例の場合、エンコーダは入力発話に対して、最終的に一つのベクトル(文脈ベクトル)を出力する。代表的なエンコーダは Recurrent Neural Network[82]と呼ばれる時系列を扱うことができるモデルであり、モデル内部に潜在的な状態(潜在状態と呼ばれ、ベクトルで表されるもの)を持っており、入力単語を用いて、その潜在状態を更新していくことができる。エンコーダは、入力発話中の単語を各タイムステップで一語ずつ読み込んでいき、ベクトルの加算または乗算で潜在状態を更新していくことで、入力発話を表す潜在状態のベクトルを作成する。

エンコーダが出力したベクトルは、デコーダと呼ばれる系列変換モデルに入力される。デコーダはエンコーダが出力した潜在状態のベクトル、および、発話生成開始の特殊トークン(EOS; End of Sentence)から出力発話「野球です」を出力する。デコーダは、各タイムステップで入力されるトークン一つに対し、出力発話に含まれる単語としての適切さを表す単語生起確率を確率分布として一つ出力する。確率分布には、単語・ベクトル変換辞書に含まれる全単語に対して、その単語がどの程度生成されそうかを表す確率値が計算されており、発話生成モデ

ルはその中で最も確率が高い単語（野球）を選択して生成する．単語生成後，次のタイムステップでは，直前に出力した単語（野球）の情報を考慮するため，それをデコーダに入力し，直前の潜在状態のベクトルと合わせて次の単語を生成する．この生成のループは発話生成終了の特殊トークン（EOS）が生成されるまで繰り返され，最終的な出力発話が決定する．

発話生成モデルを学習する際は，学習データ中の入力発話と出力発話のペアを用いて学習を行なう．入力発話に対して，学習中の発話生成モデルが出力した単語生起確率と，真の出力発話の単語生起確率の差を交差エントロピー損失（cross-entropy loss）を用いて計算し，誤差逆伝播法（back propagation）と呼ばれる手法でモデルのパラメタを更新することで学習が行われ，このパラメタの更新を学習データ中のペアを用いて繰り返すことでモデルの学習が行われる．

Vinyals らが seq2seq のモデルを提案した後，直前の入力発話のみでなく，文脈も合わせて考慮するモデルが提案された．Hierarchical Recurrent Encoder-Decoder（HRED）[76] は，入力発話をベクトルに変換するエンコーダのさらに上位に，各入力発話のベクトルを文脈のベクトルに変換するエンコーダを重ねることで，文脈を考慮するモデルである．また，Latent Variable Recurrent Encoder-Decoder（VHRED）[77] は，雑談対話における応答の多様性を捉えるために，HRED の潜在状態ベクトルにノイズを加えることで学習データ中の出力発話の多様性を吸収し，汎用的な応答の生成を抑制するモデルである．これらのモデルでは，対話における文脈や話者の意図がベクトルとして学習されるとしているが，人間による応答と比較すると単純な応答しか生成できておらず [77]，十分なモデル化ができていないとは考えにくい．

2.4.2 大量の対話データに基づく大規模発話生成モデル

近年の自然言語処理では，転移学習という機械学習手法の有効性が多数報告されている．これは，大量のテキストデータを用いて事前学習モデルと呼ばれる汎用のモデルを学習したのち，解きたいタスクのデータを用いてモデルをチューニングするという手法である．代表的なものとして，Generative Pre-training-2（GPT-2）[64]，Embeddings from Language Models（ELMO）[61]，Bidirectional Encoder Representations from Transformers（BERT）[17] などのモデルがあり，これらのモデルは自然言語処理のベンチマークタスクセットである General Language Understanding Evaluation（GLUE）[93] における高い性能が報告されている．

転移学習の枠組みは雑談対話システムにおける発話生成モデルにも応用されており，大量の対話データに基づいて，大規模な発話生成モデルを構築する研究が報告されている [1, 80]．例えば Blenderbot [80] では，まず，Reddit² から収集した大規模対話データをクリーニングし，文脈と入力発話を単一の系列にまとめ，出

²<https://files.pushshift.io/reddit/>

力発話とともに Transformer[87] と呼ばれる seq2seq と類似の発話生成モデルを学習する。その後、複数の対話データセット（個人性を含むもの、感情表現を含むもの、Wikipedia の知識を含むもの）を用いて発話生成モデルを再学習することで、14 ターンの自然な対話が実現できることを報告している。しかし課題として、プロフィール文を用いた個性の明示的な考慮 [105] や、発話における繰り返しや矛盾の抑制 [97] を試みたものの、有意な性能の改善が見られなかったことが報告されており、end-to-end での対話内容の制御は難しいと考えられる。英語で構築されている Blenderbot に対して、日本語版を作成する研究 [117] も報告されており、英語同様流暢な対話が可能なことが報告されている。誤り分析の結果として、矛盾、不適切な話題の遷移、事実関係の誤認識などが報告されている。これらの大規模発話生成モデルは雑談対話システム構築の方法として非常に発展性のある有効な手法である一方、対話の制御は難しいことが報告されているため、これらの発話生成モデルをベースとして、文脈からどう対話を進めていくかや、対話の内容をどう理解するかを検討する本研究のアプローチも必要だと考えられる。

2.5 言外の情報を扱う対話システムおよび言語処理技術

本節では、近年の対話システム研究および言語処理研究について、言外の情報に類する情報を扱った研究について説明する。

システムが聞き手となり対話を行う傾聴対話システム [126] では、ユーザ発話に対して適切な応答を行うことで対話を円滑にする試みが報告されており、これらの応答には言外の情報が含まれることから、本研究と関連がある。傾聴対話システムは対話を円滑に進めるために、さまざまな情報をユーザに伝達しうる。Inoue らは、相槌や繰り返し発話、ユーザ発話から推定した感情に基づきシステムが評価を示す発話（評価応答）など、多数の応答発話生成手法を備えた傾聴対話システム [34] を構築している。本研究で扱う発話の一部は、評価表現を含んでおり Inoue らの評価応答に位置付けられるものも存在する。下岡らは、ユーザ発話中で言及されていない必須格を尋ねる繰り返し発話や、ユーザ発話から推定した感情に基づく共感発話を生成する手法を提案している [128]。また、福田らは、Twitter から相槌として適切な発話を自動的に抽出することで、ユーザ発話に合わせて多様な相槌を行う手法を提案している [125]。

本研究は、発話に直接表れていない情報を明示的に扱うという点で、含意関係認識 [14] とも関係が深い。含意関係認識とは、テキストのペアが与えられたとき、一方のテキストの内容がもう一方のテキストの内容から推論できるか否かを判定する問題である。雑談対話から含意認識を行う研究として、Zhang らの Conversation Entailment の研究がある [101, 102]。Zhang らは、Dialogue Segment（対話の断片）と仮説（人手で作成）が与えられたとき、その真偽を判定するという課題を提案している。発話に直接表れていない情報を扱っている点は本研究に近いが、言外の情報の類型化は行っていない。Fukunaga らは、データベース検索を目的とする

タスク指向型対話システムにおいて、ユーザ発話の中でデータベースフィールドに直接対応しないが検索を行なう上で有用な情報を非明示的条件と定義し、それを抽出、分類する手法を提案している [20]. また、理解対象のテキストに対して、書かれていない情報を外部から取得するタスクとして、Document Enrichment が提案されている [103, 104]. これは、与えられた文書に対して、Wikipedia などの外部リソースを利用して、元文書と関連する追加の知識を付与するタスクである. これらの研究は、テキストからの含意認識に関する有効な手法を提案しており、言外の情報の推定にも適用できる可能性があると考えている.

近年、自然言語文に対して、クラウドソーシングを用いて言外の情報を大規模に付与する試みが報告されている. 特定の目的を達成するために必要なイベントの流れをスクリプトとして表現したスクリプト理論 [73] は、イベントに関する言外の情報を扱った代表的な研究と言える. スクリプトを大規模に収集した研究として、典型的なイベントのスクリプトを収集した研究 [95] や、クラウドワークの実体験をスクリプトとして収集した研究 [52] がある. Mostafazadeh らは、スクリプト理解のベンチマークとして代表的な Narrative Cloze Test [10] に代わるタスクとして、イベント系列のデータを収集しその結末となるイベントの正誤を判定するタスクを提案している [54]. スクリプト理論の他、人物の意図や反応に関する言外の情報を扱った研究として、Rashkin らは、入力されたイベントに関する人物の意図や反応を推定するタスクを提案し、データセットを構築している [67]. Sap らは、イベントや言外の情報をノードとして表現し、それらの間の推論の関係をエッジとして表現した知識グラフを構築している [71]. これらの研究は、自然言語文に対して言外の情報を付与している点で本研究と近いが、対話を対象としていない.

第5章で扱う議論対話は、発話に表れない命題間の関係を扱うため、言外の情報を扱う対話の一種と考えられる. 近年雑談対話システムを高度化し、論理的な対話を実現しようという試みとして、議論対話システムを構築する研究が報告されている [72, 66, 30]. Sato らは意思決定支援への応用を目的として、政治ドメインを対象に、特定の意見 (例.「カジノは経済を活性化させる」) を表す主張を入力とし、コーパスから関連する意見を抽出し、人手で整備した意見に関するフレーズ辞書に基づいて支持や反論を出力として返すシステムを構築している [72]. Rakshit らは、雑談対話システムを高度化することを目的として、議論を行なう Web フォーラムの構造化データを抽出元とし、入力された意見 (例.「死刑制度に反対である」) に対して、類似する意見を表す投稿を検索し、その意見に対する投稿をシステム発話として出力することで議論を行なうシステムを構築している [66]. Higashinaka らは、特定のトピック (例.「老後は都会よりも田舎に住むべきだ」) に関する典型的な意見、および、それに対する支持や反論の論拠を発話形式でノードとして記述した木構造を構築し、入力発話と類似する内容の知識を検索し、その発話と接続関係にあるノードを発話として返すことで議論を行なうシステムを構築している [30]. これらのシステムは既存の議論に関するリソースや人手で構築した議論の

知識を用いることで論理的な応答が可能であるが、あらかじめ用意できるリソースのカバレッジには限界があり、オープンドメインな議論を行なうことは難しいと考えられる。

対話システムとは異なるが、自然言語処理において、Argumentation Mining[44, 79]というタスクの研究が盛んに取り組まれている。これは、議論を含むテキストを入力とし、その中で議題、議論の立場、意見を表す命題を抽出し、さらにそれらの間の関係として支持や反論などを推定することで、テキストに含まれる議論を構造化する研究である。議事録から関心のある議論を自動的に構造化することができれば、議論の内容が容易に把握できるようになり、議論に関する意思決定が円滑になることが期待されている。Argumentation miningで扱われる議論の構造は、議論における典型的な支持や反論の方法論を体系的にまとめたToulminの議論構造[86]やWaltonの議論スキーム[92]を用いて表現される。Argumentation Miningは評価型ワークショップの形でも盛んに研究されているが、命題の同定や命題間の関係推定には高度な言語理解を必要とすることから、選択式の命題間の関係推定タスクでも7割程度の性能に留まっている[25]。

Argumentation Miningに関連して、命題に対して支持や反論となる命題を生成する研究が近年取り組まれ始めている。Biluらは、入力された命題に対して、noやnotなどの挿入句を命題に挿入することでその命題の否定形となる命題を自動的に生成する手法を提案している[6]。Huaらは、政治のドメインを対象に、入力された命題に対して関連するテキストをWikipediaから検索し、入力された命題と関連するテキストを用いて意見を生成する手法を提案している[33]。Parkらは、議論における発話生成のモデルが意見間の関係を一対一でしか学習できないことに着目し、多対多の関係を明示的に学習する発話生成モデルを学習することで、単一の命題に対して多様な命題を生成できることを報告している[59]。また、Hideyらは、対立意見を表す英語の定型句である「Fixed That for You (FTFY)」という表現と共に出現する投稿のペアをRedditから抽出することで、例えば「Windows XP」と「Linux」のように、対立関係にあるフレーズを発話（「Windows XPは最高のOSだ」）とともに収集、発話生成モデルとして学習することで、対立意見を生成する手法を提案している。これらの研究は主に意思決定への応用を目的として研究が進められており、対話システムへの適用が想定されていない。議論を行なう対話システムでは、システムの意図や立場に応じて、議論のトピックについて賛成なのか反対なのか、一貫した発話を返す必要があるが、先行研究では発話の立場を明示的に考慮していないため、そのままでは適用が難しいと考えられる。

2.6 本章のまとめ

本章では、言外の情報、および、対話システムに関する先行研究について説明した。まず、対話の理論研究について説明し、言外の情報はこれまで、対話の理論研究の中でその伝達の枠組みが議論されてきたものの、具体的にどのような情報

が伝わるのかは定量的に明らかにされていないことを述べた．次に，本研究で扱うオープンドメインにおけるテキストベースの雑談対話システムの位置付けを明確にするために，代表的な対話システムの類型について述べた．雑談対話システムを構築する手法として，従来のルールベースの手法と抽出ベースの手法，および，近年盛んに研究されている深層学習ベースの手法について説明し，それぞれの手法の問題点や課題について述べた．最後に，言外の情報に類する情報を扱った対話システム，および，言語処理の分野の研究について概観した．

第3章 雑談対話における言外の情報の収集と類型化

本章では、雑談対話において発話からどのような「言外の情報」が理解可能なかを明らかにする。対話において人間が発話から理解する情報のうち、発話の内容とは異なる、明示されていなくとも人間が理解する情報、これを「**言外の情報**」(perceived information)と定義し、言外の情報の収集、および、その類型化を行う。具体的な手順として、対話テキスト中の各発話に対して、複数の作業者に言外の情報を記述してもらうことで、多数の言外の情報を収集する。さらに、収集された言外の情報を用いて、複数の作業者が人手で類似する事例をまとめ上げていくことで、言外の情報の類型化を実施する。最後に、作成された類型の妥当性を評価するために、人手で言外の情報に対してアノテーションを行い、作業仲間一致率を調査する。

3.1 言外の情報の収集

雑談対話における言外の情報の類型を構築するため、まず、言外の情報のデータ収集を行った。ここでは、雑談対話における言外の情報の収集方法について述べる。収集手順として、まず、言外の情報を記述する対象となる対話テキストを用意する。次に、複数の作業者が対話テキスト中の各発話に対して言外の情報を記述することで、言外の情報を収集する。

3.1.1 言外の情報の収集対象となる対話テキスト

一般的かつ多様な言外の情報を網羅的に収集するためには、言外の情報の収集対象となるテキストの選択が重要となる。本研究では、言外の情報の収集対象として、Higashinaka らによって収集された雑談対話コーパス [27] を用いた。このコーパスには、2 名の話者がテキストチャットを行ったログが 3,680 対話収録されており、多数の話者による発話、および、様々な話題が対話中に含まれている。そのため、本研究で利用する対話テキストのコーパスとして適切である。言外の情報の記述対象として、このコーパスに含まれる 30 対話をランダムに選択した。記述対象として利用した 30 の雑談対話は 43 名の話者による 1,257 発話（1 対話あた

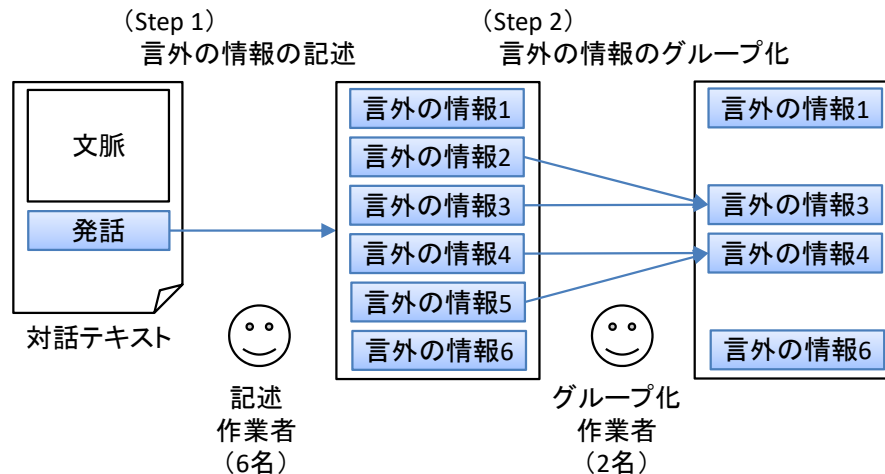


図 3.1: 対話テキストを対象とした言外の情報の収集手順

りの平均発話数は 41.9) を含んでおり、本研究の主張を支持するのに十分な規模であると考えている。

3.1.2 言外の情報の収集手順

図 3.1 に言外の情報の収集手順を示す。作業は言外の情報の記述、および、記述された言外の情報のグループ化の 2 ステップで行う。

Step 1 言外の情報の記述

6 名の記述作業者が対象となる対話に含まれる各発話に対して、言外の情報を自然文として記述する。記述作業者は対話における最初の発話から最後の発話まで順に言外の情報を記述する。特定の発話に対する言外の情報を記述する際は、その発話までの文脈のみを参照し、その発話よりも後の文脈は参照しない。

記述作業者は言外の情報を記述する際に、言外の情報として、対象となる発話の単なる言い換えの文、発話中で省略された単語を補完した文、一般常識であり、記載するに値しないと判断できる文（例えば、「人間はパンを食べられる」）を記述しないよう教示した。

言外の情報は必ず主語と述語を含むようにし、独立した文としてその文意が理解できるように記述する。また、一つの言外の情報は一つの文で記述する。各記述作業者は作業者同士で相談などはせず、独立に言外の情報を記述する。

Step 2 言外の情報のグループ化

2 名のグループ化作業者が、各発話に対して記述された言外の情報の中で、意味的に類似するものをグループ化する。この作業は、複数の作業者が記述

表 3.1: 雑談対話テキストを用いて収集された言外の情報の数

	雑談対話	言外の情報
文数（合計）	1,257	11,533
文数（異なり）	1,094	8,794
単語数（合計）	10,856	116,413
単語数（異なり）	1,596	3,740

した言外の情報の中には同じ意味内容のもの（例えば、「Aは富士山が大好きだ」と「Aは富士山が好きである」）が存在する場合があるため、それらを一つにまとめるために行う。

グループ化作業者は記述作業者とは異なる作業者で構成されるようにした。グループ化作業者は、作業者同士で相談しながら、内容語、モダリティ、時制などから内容が同じと判断できる言外の情報を一つのグループにまとめ上げる。言外の情報をグループ化した際には、グループの中から一つ最も端的に記述された言外の情報を選択し、そのグループを表す言外の情報とする。

上記の手順にならって、各対話テキストに対して6名の記述作業者がStep 1の記述作業を行い、2名の作業者がStep 2のグループ化作業を実施することで、言外の情報を収集した。記述作業では、一人あたりの作業量を少なくするため、のべ12名の記述作業者に作業を依頼し、そのうち6名が対象となる対話テキストの半数に記述作業を行い、残りの6名が残りの半数の対話テキストに対して記述作業を行った。一般的な言外の情報を幅広く収集するため、記述作業者は言語学の専門家などを除いた一般的な作業者になるようにし、バックグラウンドや性別が偏らないようにした。実際に作業者に提示した教示は付録3.5.1に示す。

3.1.3 収集された言外の情報

表 3.1 に対象の雑談対話、および、収集された言外の情報における文数と単語数を示す。作業対象の対話に含まれる1,257発話に含まれる発話に対して、Step 1の記述作業で12,723文の言外の情報が記述され、Step 2のグループ化作業で11,533文（91%）の言外の情報にグループ化された。言外の情報の1文あたりの平均単語数は10.1単語であった。雑談対話と言外の情報について、文数、および、単語数の合計を比較すると、雑談対話のおよそ10倍の文数、および、単語数が言外の情報として収集されており、発話の表層と比較して多くの情報が発話から言外の情報として理解できることがわかる。

図 3.2 に記述対象の雑談対話テキスト、および、特定の1発話に対して記述された言外の情報の例を示す。雑談対話テキストには2名の話者AとBによる発話が37発話（ U_1 から U_{37} ）収録されており、各発話に対して言外の情報が記述されて

言外の情報の収集に利用する雑談対話テキスト

U_i	話者: 発話	
U_1	A: こんにちは！	
U_2	B: こんにちは。	
U_3	A: 秋が近付いてきましたね。	
U_4	B: 私もそう思います。	
U_5	B: せみも静かになってきましたね。	
U_6	A: 鈴虫の鳴き声に変わりましたよね。	
U_7	B: 紅葉が楽しみです	
U_8	A: 紅葉を見るのが好きなんですか？	U_{13} に対して収集された言外の情報 P_1 : B は遠出をいとわない P_2 : B は行動的だ P_3 : B は気分屋だ P_4 : B はアウトドア派だ P_5 : B は山が好きだ P_6 : B は富士山が好きだ P_7 : B は富士山の紅葉が好きだ P_8 : B はアウトドアが好きだ P_9 : B は関東に住んでいる P_{10} : B は富士山の近くに住んでいる P_{11} : B は A に驚いてもらいたい P_{12} : 富士山は紅葉が有名だ
U_9	B: 好きなんです！	
U_{10}	B: というか、山とか行ってついでに温泉入ったら...と思うと余計楽しみになってきます。	
U_{11}	A: それはいいですね！	
U_{12}	A: 毎年どこか紅葉を見に行かれたりするんですか？	
U_{13}	B: 気が向いたら、富士山に行ったりしますね。	
...		
U_{36}	A: また次の機会に話しましょう！	
U_{37}	B: よろしく願います。	

図 3.2: 収集対象の雑談対話テキストと収集された言外の情報の例

いる．言外の情報は B の発話 U_{13} に対して記述されており，6 名の作業者が一つ以上の言外の情報を記述した結果，12 文の言外の情報（ P_1 から P_{12} ）が収集されている．

対話テキストの内容から，話者 A と B が紅葉の話題について話しており， B が紅葉を見に富士山に行くという内容が理解できる． U_{13} の発話から， B が毎年富士山に行くことがわかるため， $P_1 \sim P_4$ のように，行動的な性格の可能性があるという言外の情報が記述されている．また，新しく富士山の話題が出てきており， B が毎年富士山に行くという事実から， $P_5 \sim P_8$ のように， B が山や富士山のことを好きという言外の情報が記述されている． P_9 と P_{10} については，毎年富士山に行くということから， B が富士山の近くに住んでいる可能性が高いと考えられるために記述されている． P_{11} については，紅葉を見るために毎年行くとは一般的に考えづらい富士山について B が発話していることから， A を驚かせて話を盛り上げたいと作業者が考えたために記述されたと考えられる． P_{12} については，毎年紅葉を見に行ける場所は，一般的に紅葉が有名だと考えることができるため，記述されている．これらの例から，話者 B の性格や普段の行動，富士山についての情報など，様々な情報が言外の情報として伝達されていることがわかる．

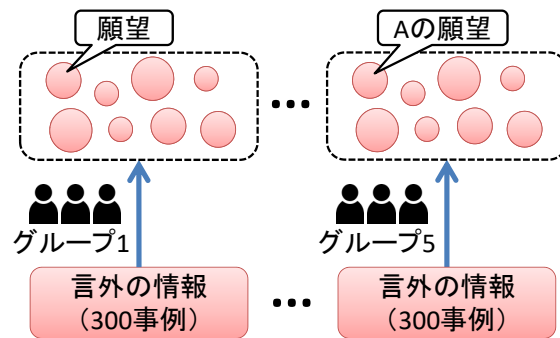


図 3.3: 言外の情報の類型化手順 (Step 1. 言外の情報のまとめ上げ)

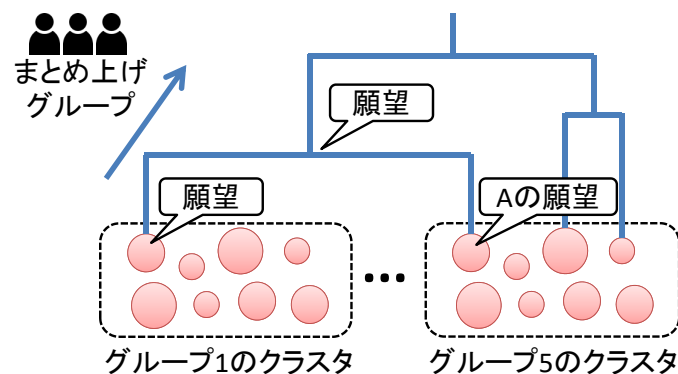


図 3.4: 言外の情報の類型化手順 (Step 2. まとめ上げ結果の統合作業)

3.2 言外の情報の類型化

収集した言外の情報事例を用いて、言外情報の類型化を行った。類型を作成するには、特定の作業員による恣意的な類型ではなく、一般性を持った類型が作成されるよう、複数の作業員が類型化を行い、それらを統合するようにした。異なる複数の類型を用いて、その共通部分をまとめ上げることで、全ての作業員の意見が反映された、一般性の高い類型が作成できると考えた。なお、以降説明する類型化の前段階の試みとして、対話理論に基づいて発話と言外情報の関係を類型化する試みを実施したが、類型の改善を繰り返しても、アノテーションの一致率が上がらないという問題が発生した。そのため、本研究では、言外情報をまとめ上げていく形で類型を作成した。対話理論に基づく類型化の試みの詳細は、付録 3.5.3 で説明する。

3.2.1 言外情報の類型化手順

図 3.3、および、図 3.4 に言外情報の類型化手順を示す。類型化作業は、言外情報のまとめ上げ、および、まとめ上げ結果の統合作業の二つの手順で行われる。

Step 1 言外の情報のまとめ上げ

3名の作業員からなるグループが、与えられた言外の情報の中で類似する事例を手で相談しながらまとめ上げ、言外の情報の類型を作成する。与えられる言外の情報は、データ収集手順の Step 2 で作成したグループの代表を表す言外の情報のみである。類型の粒度は次の統合作業で調整されるため、作業員には言外の情報の類似性を判断する基準は与えず、個々のグループで基準を決めてまとめ上げ作業を行う。

作業は最も類似すると考えられる言外の情報から順にグループ化を行っていく、全ての言外の情報がいずれかのグループに含まれるようになるまで作業を行う。作成されるグループは要素として言外の情報の事例を持ち、階層構造（グループの入れ子）にはしない。言外の情報のグループには、グループの特徴を表す名前（ラベル）を付与するようにする。この作業を異なる複数のグループが行う。

Step 2 まとめ上げ結果の統合作業

複数の作業員グループによって作成された異なる言外の情報のグループを、Step 1 とは異なる作業員で構成される統合作業員グループが統合する。統合作業員グループは、作成された全ての言外の情報のグループを確認し、階層的に統合する。

具体的には、最も類似する言外の情報のグループを二つ決定し、それらを統合することで新しいグループを作成する。Step 1 で言外の情報のグループに付与されたラベルは類似する言外の情報のグループを選定する際に利用される。この統合作業を繰り返し行い、最終的に言外の情報グループの数が一つになるまで繰り返す。Step 1 と同様に、統合で新しく作成された言外の情報のグループにはラベルを付与する。

収集された言外の情報は1万事例を超えており、これら全てをまとめ上げることはコストが高いため、事例のサンプリングを行った。具体的には、収集に利用した雑談対話テキストの中から、ランダムに5対話を選択し、対話テキストごとに300事例の言外の情報をランダムに抽出し、計1,500事例（=300事例×5対話）を類型化のために利用した。

Step 1 のまとめ上げ作業員グループとして、異なる5組のグループが、それぞれ300事例の言外の情報に対して作業を行った。Step 2 の統合作業員グループは Step 1 のまとめ上げ作業員グループとは異なる作業員になるようにし、異なる5組のまとめ上げ作業員グループが作成した言外の情報のグループを統合した。Step 1 のまとめ上げ作業は類似する言外の情報をまとめる作業であり、常識的な言語処理能力を有していれば専門家でなくとも作業が可能と考えた。そのため、専門家であることを条件とせずに作業員を募集した。結果的に、作業員に言語学の専門家は含まれなかった。また、作業員の男女比が偏らないようにした。Step 2 の統

合作業者では確認する事例が多く、またその組み合わせも多数存在するため、作業員としてアノテーション作業の専門家を割り当てた。作業員に提示した実際の教示は付録 3.5.2 に示す。

3.2.2 言外の情報の類型

表 3.2 に、本作業で作成された言外の情報の類型を示す。類型化作業はボトムアップに行われるため、階層的な類型になっており、表には最も上位の 4 階層の類型が示されている。レベルは類型の階層に対応しており、最も左の列（レベル 1）が最も上の階層を表し、右の列から左の列に向かって類型が統合されている。表から、言外の情報は大きく 2 つの種類、思考内容と事実に分けられることがわかる。思考内容は話者の考え、および、願望で構成されており、事実は話者に関する事実である話者の事実、および、話者に直接言及していないその他の事実で構成される。

各ラベルの出現頻度の割合に着目すると、レベル 1 については、思考内容が 55%、事実が 45% とおよそ半数ずつ出現しており、思考内容の方が多く出現していることから、雑談では思考内容が最も多く言外の情報として伝わるということがわかる。レベル 2 については、話者の事実が 38%、信念が 36% と多く出現しており、合わせておよそ 70% が話者に関連する考えと事実で占められていることがわかる。また、願望が 19% と 3 番目に多く出現しており、多くの願望が話者間で伝わっていることがわかる。話者とは直接関係のない事実的な内容（例えばニュースなど）は全体のうち 7% 程度しか出現しておらず、雑談対話で理解すべき言外の情報は話者と直接関係のある内容が 9 割以上であることがわかる。レベル 3 については、信念 1 が 31% と最も多く出現しており、言外の情報の中で最も多いのは話者自身の考えであることがわかる。

付録 3.5.4 節の表 3.7 に、類型化の過程で作成された全てのラベルとその割合を示す。表には 8 つの階層と、80 個の類型が示されており、表 3.2 は 6 から 8 の階層に対応する。5 組の作業員グループが独立に作成した類型である 1 の階層には、表 3.2 に示される類型と類似するラベルが見られることから、5 つの作業員グループの結果は粒度こそ異なるものの同様の類型になっていることが確認できる。

以降、レベル 3 の類型の各ラベルの定義、および、各ラベルで代表的な言外の情報の種類とその具体例を示す。なお、レベル 4 の類型は、異なる上位ラベルを持ちながら同じ内容を表す言外の情報が含まれるケースが見られたため、本研究では、レベル 3 の類型を最も詳細な言外の情報の類型とした。

表 3.2: 雑談対話における言外の情報の類型. 括弧内の数値は対応する言外の情報
 が出現する頻度の割合を示す. 言外の情報をボトムアップにまとめていくこ
 とで類型を作成したため, 階層的な類型になっている. レベルは類型の階層を表
 しており, レベル1が最も上位の階層を表す.

レベル1	レベル2	レベル3	レベル4
思考内容 (55.1%)	考え (35.8%)	信念1(30.7%)	思っている・考えていること (1.6%)
			話者は... と思っている (1.0%)
			趣向 (13.9%)
			対象に対する印象・評価 (6.7%)
			感情 (7.5%)
		信念2 (5.1%)	相手に関する印象 (4.4%)
			話者と聞き手の関係性 (0.7%)
	願望 (19.3%)	願望1 (9.9%)	聞き手を意識した自分の願望 (3.1%)
			自己完結の願望 (3.2%)
			自分の願望 (3.5%)
		願望2 (9.4%)	相手にして欲しいこと (1.1%)
			聞き手に対する願望 (8.3%)
事実 (44.9%)	話者の 事実 (37.9%)	属性 (20.2%)	話者の特徴 (19.5%)
			所有しているもの
		行動 (14.4%)	話者の過去・経験 (8.9%)
			話者の現在の状態 (5.5%)
	その他の 事実 (7.0%)	環境 (3.3%)	話者と話者の周辺の状態 (1.6%)
			話者の周辺の状態 (1.7%)
		事実1 (3.9%)	ものに関する事実 (0.7%)
			一般的なこと (0.3%)
			客観的な事実 (1.7%)
			一般的な事実 (1.1%)
		事実2 (3.1%)	不確定な事実 (1.2%)
			事実 (1.4%)
			起きる可能性があること (0.5%)

信念 1, 2 話者の信念に対応しており、信念 1 は話し相手以外に対する信念、信念 2 は話し相手に対する信念に対応する。信念 1 は、話者の考え、趣向、感情を含む。

- (1) **考え**：「A は東京の物価に不満がある」、「A は日本が安全だと思っている」
- (2) **趣向**：「A はゲームが好きだ」、「A は旅行に興味がある」
- (3) **感情**：「A は興奮している」、「A は B に褒められて嬉しい」

信念 2 は、話し相手の発言や話し相手自身に対する信念を含む。

- (4) **相手の発言に対する信念**：「A は B に共感した」、「A は B の話を信じられない」
- (5) **相手自身に対する信念**：「A は B を心配している」、「A は B をすごいと思った」

願望 1, 2 話者が持つ願望に対応しており、願望 1 は話者の主体的な願望、願望 2 は話し相手が関連する願望に対応する。願望 1 は、話者のみが関係する願望と、相手も関連のある願望を含む。

- (6) **話者のみ関係がある願望**：「A は富士山に行ってみたい」、「A は早く夏が終わってほしい」
- (7) **相手も関係がある願望**：「A は話題を変えたい」、「A は B に趣味の話がしたい」

願望 2 は話し相手が行動することで達成される願望や、話者と話し相手の両方が行動することで達成される願望を含む。

- (8) **相手に対する願望**：「A は B に褒めてもらいたい」、「A は B の趣味を知りたい」
- (9) **相手と共に成し遂げる願望**：「A は B と話を合わせたい」、「A は B に映画の魅力を伝えたい」

属性 話者のユーザ属性に対応し、知識・能力、プロフィール、性格を含む。

- (10) **知識・能力**：「A はロードバイクについて詳しい」、「A はお酒が飲める」
- (11) **プロフィール**：「A は女性だ」、「A は結婚している」、「A は東京に住んでいる」
- (12) **性格**：「A は真面目だ」、「A は意思が強い」

行動 話者の行動に対応し、習慣、過去・経験、対話中の様子を含む。

- (13) **習慣**：「Aはテレビをよく見る」,「Aはあまり外に出かけない」,「Aは車を運転する」
- (14) **過去・経験**：「Aは昨日両親と話した」,
「Aは北海道に行ったことがある」
- (15) **対話中の様子**：「Aは話題を変えようとしている」,
「Aは得意気だ」,「Aは次に言うことを考えている」

周辺環境 話者を含む周辺の環境に対応し、人間関係や生活環境を含む。

- (16) **人間関係**：「Aは家族と仲が良い」,「Aは友達が多い」,「Aの夫はテレビをよく見る」
- (17) **生活環境**：「Aの仕事は転勤が多い」,「Aの自宅にはクーラーがある」,
「Aの実家は関東にある」

事実 1, 2 話者と無関係な客観的事実に対応し、確定的な事実と不確定な事実を含む。

- (18) **確定的な事実**：「今年は猛暑が続いている」,「富士山は紅葉が有名だ」
- (19) **不確定な事実**：「農作物が不作になりそうだ」,「韓国旅行はリーズナブルだ」

作成された類型を見ると、対話システムの内面のモデルとしてよく検討される Belief-Desire-Intention モデル [7] で扱われる信念や欲求（願望）に関するものが含まれていることがわかる。また、ユーザモデルの研究でよく検討される、ユーザ属性などのプロフィール情報 [31] が含まれている。このことは、ユーザに関して一般に理解すべきとされている情報が雑談対話においても重要であることを示唆するが、これら以外の情報として、話者の行動や環境といったことも重要であるということは興味深い。雑談対話においては、個人のエピソードや社会関係も重要であることを反映していると考えられる。雑談対話システムがユーザ発話を理解する際には、これらのことも考慮する必要がある。

3.3 類型の評価

作成された類型を評価するために、類型の作成とは無関係な3名の作業者が言外の情報の類型のアノテーションを行った。アノテーションには類型のレベル3を使用した。各作業員（アノテータ）は、類型の作成時に使用していない対話の中からサンプリングした10対話に対して記述された言外の情報3,000事例に対して

表 3.3: 言外の情報に対する類型をアノテーションした際のアノテータ間一致率. レベルは言外の情報の類型における階層を表しており, レベル3はアノテーション時に用いた類型における一致率, レベル1とレベル2はレベル3の類型をまとめた際の一致率を表す.

	レベル1	レベル2	レベル3
単純一致率	0.90	0.86	0.75
Fleiss' κ	0.80	0.79	0.69

アノテーションを行った. アノテーションを行う際は, 言外の情報のみを見て類型を付与するようにし, 言外の情報を収集する際に利用した対話の文脈, および, 言外の情報の記述の対象となった発話は参照しないようにした.

表 3.3 に言外の情報の類型をアノテーションした際の一致率として, 単純一致率, および, Fleiss の κ 値を示す. レベル3に示されている値は実際にアノテーションした類型における一致率を表しており, レベル2, および, レベル1はレベル3の類型をまとめた際の一致率を表す. κ 値は0.69から0.80となっており, 高い一致率で類型が付与されていることがわかる.

今回のアノテーションでは, 類型に「その他」や優先度がないにもかかわらず高い一致率を示していることから, 類型がカバーする言外の情報に漏れ(どの類型にも属さない事例)や重複(複数の類型が付与可能な事例)が少ないと考えられる. また, 異なる作業員でも類型の理解が容易であることがわかる. これらの事実は, 類型が必要な概念を正しく分類できていることを示しており, このことから我々は本類型が妥当なものであり, 一般性がある程度担保できているものと考えている.

表 3.4 に, 付与された類型の混同行列を示す. 表には, 2名のアノテータの組み合わせから得られる混同行列の合計が事例数で示されている. この表から, まず, 信念1と信念2, 願望1と願望2の間で揺れが生じていることがわかる. これは, どこまでの言外の情報を話者自身に関するものとするかが作業員によって揺れているためだと考えられる. 例えば, 信念1と信念2で揺れる例として, 「AはBの言う紅茶に関心がある」という言外の情報が見られ, 紅茶自体に関心があると捉える場合(信念1)と, Bが言及した紅茶として関心の対象にBも含める場合(信念2)で揺れる.

また, 信念1と願望1, 属性, 行動が揺れやすいことがわかる. 揺れが生じた事例を確認したところ, 信念1と願望1については, ユーザの信念に基づく願望, 例えば「Aは綺麗でいたい」や「Aは情熱的に生きたい」という言外の情報が見られた. 信念1と属性については, ユーザの趣味や思考に従って行う行動, 例えば「Aは自然思考だ」や「Aは冒険好きだ」という言外の情報が見られた. 信念1と行動については, 信念に基づく対話中の行動, 例えば「Aは恥ずかしがっている」や「Aは話題がいきなり変わって戸惑っている」という言外の情報が見られた.

表 3.4: 言外の情報 3,000 件に対して類型（レベル 3）をアノテーションした際の混同行列. 3 名の作業結果における混同行列の合計が示されている. 下線は類型の不一致が顕著な事例（100 事例以上のもの）を表す.

	信念 1	信念 2	願望 1	願望 2	属性	行動	周辺 環境	事実 1	事実 2
信念 1	2,076	<u>152</u>	47	4	<u>142</u>	<u>117</u>	10	7	4
信念 2	32	452	6	4	4	54	0	0	0
願望 1	<u>142</u>	40	633	<u>201</u>	1	31	0	2	2
願望 2	6	42	80	404	0	4	0	0	0
属性	<u>233</u>	13	0	0	1,302	87	35	9	2
行動	<u>175</u>	82	8	1	<u>171</u>	1,305	41	16	2
環境	14	0	0	0	13	9	139	8	0
事実 1	4	1	0	0	2	8	29	225	59
事実 2	7	0	0	0	0	0	5	97	154

その他に、属性と行動にも揺れが生じやすいことがわかる。事例を調査したところ、話者の行動から得られるユーザ属性、例えば、「A は大学に通っていた」や「A は定期的に運動をしていない」という言外の情報が見られた。

これらの揺れから、一部の言外の情報は複数の種類の情報（例えば、属性かつ行動）として捉えることができる場合があることがわかる。この結果から、システムが言外の情報を扱う際には、マルチレベルの枠組み、例えば、特定の言外の情報をいずれか一つの類型に分類するのではなく、話者の自分自身に関する信念でもあり、願望でもあるという理解、も検討したほうがよい可能性がある。

3.4 本章のまとめ

本研究では、雑談対話において発話から理解すべき情報（言外の情報）の類型化を実施した。雑談対話における各発話の言外の情報を作業者が記述することで言外の情報を収集し、複数の作業者が類似する言外の情報をクラスタリングすることで階層的な類型を作成した。

作成された類型は言外の情報の類型をアノテーションした際のアノテータ間一致率で評価され、最も詳細な類型の一致率は κ 値で 0.69 となり、作成した類型が妥当であることが確認された。本研究は我々の知る限り、雑談対話における言外の情報の類型化を試みた最初の研究である。得られた知見は将来的に、発話の内容を越えて明示的に表れない重要な情報を理解し対話を行うシステムにおいて、言外の情報の推定手法や利用方法を検討する際に有用である。

今後は、言外の情報を対話から自動的に推定する手法の研究を進め、推定した言外の情報を基に適切な発話を生成する、適切な行動を選択する対話システムを構築するなどの課題に取り組む予定である。また、異なる作業者が類型化を行っても本論文で示す類型化にたどり着くかどうかなど、類型の一般性を高めるために、より多くの対話データや作業者をを用いて類型を評価していくことも重要な課題である。

3.5 本章の付録

3.5.1 言外の情報を記述する作業者への教示

2名の人物（AとB）の対話テキストを読み、発話に表れていないけれども、「話し手が伝えたい情報や聞き手に伝わる情報」（言外の情報）を記述してください。言外の情報には、例えば、人物の属性（例. 既婚かどうか）などの事実、話し手や聞き手が「～である」と思っていること（信念）、「～したい」と思っていること（欲求）、「～するつもりだ」と思っていること（意図）などがあります。以下に例を示します（発話：「」、言外の情報：『』，S：話し手，H：聞き手）。

- S: 「昨日は妻と出かけました」（昨日何をしていたかという話題にて）
 - － 『Sは男性である』（妻がいるため）
 - － 『Sは結婚している』（「妻がいるため」）
 - － 『Sは妻と仲がいい』（妻と一緒に出かけているため）
 - － 『Sは妻とのお出かけについて話したい』（自ら出かけたことを話しているため）
- S: 「すごいですね！」（ダイエットが成功したという話題にて）
 - － 『ダイエットは大変である』（「すごい」と反応しているため）
 - － 『SはHを尊敬している』（「すごい」と反応しているため）
 - － 『SはHを喜ばせたい』（褒めているため）

言外の情報かどうかは、(Sは)「～と伝えた」もしくは(Hに)「～と伝わった」に当てはめることができるかどうかで判断してください。また、言外の情報と発話を照らし合わせて、言外の情報がほとんどの場合に成り立つと考えられれば、言外の情報として記述してください。

- S: 「すごいですね！」
 - － 『SはHを尊敬している』（OK）

- * 「SはHを尊敬していると伝えた」と当てはめられる
- * 「すごい」と発言しているため、尊敬していると考えられる
- － 『すごいH』(NG)
 - * 「SはすごいHと伝えた」は不自然であり、当てはめることができない
 - * 「すごい」と発言しているため、内容に問題はなく、成り立っている
- S: 「昨日は妻と出かけました」
 - － 『Sは結婚している』(OK)
 - * 「Sは結婚していると伝わった」と当てはめられる
 - － 『Sは妻と仲が悪い』(NG)
 - * 「Sは妻と仲が悪いと伝わった」と当てはめられる
 - * 妻と一緒に出かけしており、仲が悪いとは言っていないため、成り立たない

言外の情報を記述する際には、以下に注意をしてください。

- 2名の人物について言及する場合は、AかBで記述してください。
 - － 『Aはピアノを習っている』(OK: Aを用いて言及しているため)
 - － 『話し手はピアノを習っている』(NG: AやBではなく「話し手」を用いて言及しているため)
 - － 『ピアノを習っている』(NG: Aが省略されており、主語が何かわからないため)
- 発話と同じ意味や近い意味を持つ言外の情報を書かないようにしてください。
 - － S: 「昨日妻と出かけました」→『Sは昨日外出した』(NG)
 - － S: 「すごいですね」→『SはHを褒めている』(NG)
 - － S: 「そうですね。暑いですね。」→『SはHに同意している』(NG)
- 同じ意味や近い意味を持つ言外の情報を複数書かないようにしてください。
 - － 『Sはピアノができる』, 『Sはピアノが好きだ』, 『Sはピアノが得意だ』, 『Sはピアノに自信がある』(NG)
 - － 『Sはピアノができる』, 『Sはピアノを持っている』, 『Sはピアノを習っている』(OK)

- 言外の情報はなるべく短い1文で記述し、複数の言外の情報がある場合は、複数の文に分けて書いてください。
 - － 『Sは結婚している』(OK)
 - － 『Sは結婚している』, 『Sは男性だ』, 『Sはピアノが好きだ』(OK)
 - － 『Sは結婚しており、男性であり、ピアノが好きだ』(NG)
- 文には必ず主語と述語(「何」が「どうした」か)が含まれるようにしてください。
 - － 『Sは結婚している』(OK: 主語と述語があるため)
 - － 『結婚している』(NG: 主語が省略されているため)
 - － 『Sの結婚』(NG: 主語と述語が不明)
 - － 『結婚』(NG: 主語と述語が不明)
- 原則1発話ごとに1文以上書く必要がありますが、言外の情報がないと考えた場合は、無理に書く必要はありません。
- 1つの発話に対して、矛盾する言外の情報を書かないでください。ただし、話の展開などによって、別の発話に対して記述した言外の情報と矛盾することは構いません。
 - － A: 「ピアノが好きなんです」
 - － B: 「そうなんですか」→『Bはピアノが好きではない』(OK)
 - * 返事がそっけなく、関心があるように見えないため
 - － A: 「Bさんは楽器とかやるんですか」
 - － B: 「実は自分もピアノを弾きます」→『Bはピアノが好きだ』(OK)
 - * 前の言外の情報と矛盾しているが、話の展開で判明したことなので問題ないため
- 人物は便宜上AとBで表されていますが、各対話において、異なる別の人物を表します。
- 一部の対話は、別の対話の続きになっているものがあります。その場合でも、他の対話テキストと同じように、言外の情報を記述してください。
- 対話の最初のあいさつや最後の終了のあいさつは、作業対象外としています。言外の情報を書く必要はありません。
- 作業は以下の先頭の発話から順に行い、発話より後ろは見ないで言外の情報を考えてください。

3.5.2 言外の情報を類型化する作業者への教示

Step 1. 言外の情報のまとめ上げ

似ている言外の情報を同じグループにまとめて、10 個程度（最大 20 個程度まで）のグループを作ってください。表現や単語が似ているかどうかでまとめるのではなく、以下に示す例のように、言外の情報が示す内容の種類が似ているかどうかでまとめてください。なお、これらの例は一例であり、このようにしなくてはならないという例ではありません。

- 例 1: 相手への印象を表している点において似ている言外の情報
 - － 「A は B にアウトドアなイメージがある」
 - － 「A は B のことをアクティブだと思っている」
- 例 2: 相手への願望を表している点において似ている言外の情報
 - － 「A は B に旅行してもらいたい」
 - － 「A は B がジムを続けることを望んでいる」

グループ化を行う際には、各グループに、そのグループの特徴を表すラベル（例. 「○○が似ているグループ」など）、および、作業の際のコメント（例. 「○○をどのグループにするか迷ったが、△△の理由で□□にした。」など）を残してください。

3 名が並行して上記の作業を行った後、結果をすり合わせるにより、最終的なグループ化の結果を決めてください。結果をすり合わせる際には、誰か一人の結果をそのまま採用するのではなく、全員の作業結果を考慮した上で、グループ化の基準が明確、かつ、第三者が見ても妥当と思われるような結果にしてください。

Step 2. まとめ上げ結果の統合作業

5 組の作業者によって作成されたグループ化の結果を 1 つのグループになるまで統合してください。作業の流れを以下に示します。

- (1) 最も似ている言外の情報を含むと考えられるグループを 2 つ選び、1 つのグループに統合する。
- (2) 統合したグループ（統合グループ）を 1 つのグループとみなして、最も似ている言外の情報を含むと考えられるグループを 2 つ選び、1 つのグループに統合する。
- (3) 以降、上記の統合作業をグループが 1 つになるまで行う。

誰か一人の意見をそのまま採用するのではなく、全員の意見を考慮した上で、グループを統合する基準が明確、かつ、第三者が見ても妥当と思われるような統合グループを作成してください。

クラス	定義
会話の含み	婉曲的な発話を用いて伝わる言外の情報
プラン認識	発話の内容に対するプラン認識によって伝わる前提もしくは帰結となる言外の情報
談話プラン	話題の変更など、談話に関する話者の意図が伝達される言外の情報
共通認識	発話内容の理解や修正を目的としたメタ対話から伝わる言外の情報
発話のプラン認識	発話する行為に対するプラン認識によって伝わる前提もしくは帰結となる言外の情報
ユーザ属性	発話からユーザのプロフィールに関する情報が伝わる言外の情報

表 3.5: 対話理論に基づく雑談対話における言外の情報のトップダウン型の類型

3.5.3 対話理論に基づくトップダウン型の類型化の試み

本研究では言外の情報を類型化するための最初のアプローチとして、言外の情報に関する対話理論の知見を活用するために、対話理論に基づいたトップダウンの類型化を試みた。最終的に提案した表 3.2 の類型とは異なり、ここでは、対話理論が言外の情報が伝達される枠組みについて議論しているということを鑑みて、発話媒介行為、協調の原理、プラン認識、談話プラン、共通基盤などで構成される類型をあらかじめ用意し、発話と言外の情報のペアをこれらの類型に分類することを試みた。

表 3.5 に、対話理論に基づく雑談対話における言外の情報のトップダウン型の類型を示す。本類型は、筆者を除く複数人の作業者が分類を行い、その分類結果に基づいて、一致率を高めるよう類型を修正するプロセスを複数回繰り返すことで作成されたものである。類型は大きく六つの類型に分かれており、内二つの類型については内部でさらに二つの類型に分かれることから、類型の数は合計で八つとなる。以降、各類型について説明する。

会話の含み この類型は、発話の中で婉曲的な表現が用いられており、グライスの格率を違反することで言外の情報が会話の含みとして伝わるものである [22].

- (1) A: 「食べることにしては好きです」 → 言外の情報: 「A は料理が好きではない」

プラン認識 この類型は、言外の情報がプラン認識によって言外の情報として推論され、発話の前提や帰結を表すものである [2]. プラン認識の中では、decomposition のように前提や帰結以外の類型も定義されているが、提案する類型を作成する過程において、前提または帰結のどちらかのみに分類するように修正した。この類型の中で、前提と帰結は異なる類型として扱われる。

(2) A:「昨日夫と出かけました」 → 言外の情報:「Aは結婚している」
(前提)

(3) A:「休日はいつもジムに行きます」 → 言外の情報:「Aは健康だ」(帰結)

談話プラン この類型は、話題の変更など、談話に関する話者の非明示的な意図が言外の情報として伝わったものである。この類型は談話プラン[45]に関連する類型である。

(4) A:「実は私北海道出身なんです」 → 言外の情報:「Aは出身地について話したい」

共通基盤 この類型はメタ対話を通じて、言外の情報が伝達されることで共通基盤が形成されるものである[12]。例えば理解できなかった発話の内容を確認したり、発話した内容を遡って修正したりする場合がある。

(5) A:「確かにそうですね」 → 言外の情報:「AはBが言った内容を理解した」

発話のプラン認識 この類型は「プラン認識」の類型と類似しているが、「プラン認識」の類型が発話の内容についての前提や帰結であるのに対し、この類型は発話をする行為自体に対するプラン認識を表す。より正確には、この類型は話者が発話する行為の結果として、話者が聞き手に対してその行為の前提や帰結を言外の情報として推論することを期待するものであり、発語媒介行為と関連する類型であり、例えば聞き手の説得、対話の誘導などに関連する発話および言外の情報が含まれる[3]。前提と帰結は「プラン認識」の類型同様、前提と帰結は異なる類型として扱われる。

(6) A:「スキーでパラレルターンはできますか？」 → 言外の情報:「Aはスキーについて詳しい」(前提)

(7) A:「簡単な料理であれば毎日やっています」 → 言外の情報:「Aは褒めてほしい」(帰結)

ユーザ属性 この類型は発話からユーザのプロフィールに関する情報が言外の情報として伝達されるものである。

(8) A:「いつも走っています」 → 言外の情報:「Aは勤勉な性格だ」

上記の類型の妥当性を評価するために、収集した言外の情報を用いて、発話と言外の情報のペアに対して類型をアノテーションした。アノテーションの際は、対象の発話までの文脈、および、対象となる言外の情報に対して類型を一つ付与す

表 3.6: トップダウンの類型を用いた言外の情報のアノテーション結果

類型	カッパ値	頻度
会話の含み	0.23	9%
プラン認識 (前提)	0.37	18%
プラン認識 (帰結)	0.33	7%
談話プラン	0.64	12%
共通基盤	0.37	13%
発話のプラン認識 (前提)	0.58	14%
発話のプラン認識 (帰結)	0.45	9%
ユーザ属性	0.37	18%
全体	0.42	100%

るようにした。作業者には上記の類型の定義、および、数十個の練習用の事例にアノテーションを行った後、本データに対するアノテーション作業を行った。筆者を除く二名のアノテータが 150 ペアの発話と言外の情報に対して作業を行った結果、アノテーションの一致率は Fleiss のカッパ値 [19] で 0.42 となり、ある程度の一致が見られたものの、高い一致率の類型を作成することはできなかった。数回の仕様の変更の後、カッパ値の改善が見られなかったため、最終的に、対話理論に基づくトップダウンの類型化ではなく、表 3.2 で述べたボトムアップ型の類型化を試みた。

表 3.6 に、トップダウンの類型を用いた言外の情報のアノテーション結果を示す。表には各類型のカッパ値と頻度が示されている。カッパ値で示されている通り、特定の類型は他の類型と比較して高い一致率を示している。「談話プラン」と「発話のプラン認識」の類型は、他の類型と比較して一致率が高い。これは、アノテーションされた事例を分析した結果、「談話プラン」の類型はそのほとんどが話題の変更に関する言外の情報であり、また、「発話のプラン認識」の類型はその他の話者の意図に関する言外の情報であったためと考えられる。これら以外のクラスはいずれも低い一致率となっており、この原因として、各類型の定義が主観的であることが考えられる。例えば、「会話の含み」の類型に関して、会話の含みを感じるかどうかは作業者の背景や内省によって個人差があるため、言語的な場合分けなど何らかの統一的な分類規則を設けない限り、アノテーションを一致させることは難しいことが示唆される。各類型がアノテーションされた割合に着目すると、各類型は概ね同程度出現していることから、雑談対話では対話理論で議論されてきたさまざまな方法で言外の情報が伝達されることが確認できる。

3.5.4 言外の情報を類型化する過程で作成された全ての類型

表 3.7: 言外の情報を類型化する過程で作成された全ての類型. 表 3.2 に示す最終的な類型の名称は, 表中の名称に基づいて著者が作成したものである.

	Step 8	Step 7	Step 6	Step 5	Step 4	Step 3	Step 2	Step 1
1				(11) [0.7%] #物に	(3) [0.2%] Tある物の関係の特徴			
2				対する事実	(8) [0.5%] Tある物の性質を表している			
3				(5) [0.3%] P:客観的事実1(一般的なこと)				
4			(58) [3.9%] #客観的事実	(25) [1.7%] Q:客観的事実				
5				(17) [1.1%] S:その他の事実(世間の話)				
6				(18) [1.2%] #不確定事実	(12) [0.8%] #一般論	(7) [0.5%] Q:一般論		
7			(47) [3.1%] #客観的事実(不確定含む)	(21) [1.4%] R:事実	(5) [0.3%] R:世事への感情・印象			
8				(8) [0.5%] Tある物の選択可能性を表している	(6) [0.4%] Q:予想・推測			
9				(24) [1.6%] #AとAの周辺の状態	(21) [1.4%] #家族関係	(7) [0.5%] Q:AとAの家族の情報		
10				(26) [1.7%] #Aの周辺の状態	(14) [0.9%] S:Aの人間関係	(14) [0.9%] S:Aの人間関係		
11			(50) [3.3%] #背景	(23) [1.5%] Q:主語が「Aの…」の文	(3) [0.2%] S:存在			
12				(134) [8.9%] #Aの過去・経験	(130) [8.7%] #行動に基づいたAの背景	(64) [4.3%] #習慣・経験	(43) [2.9%] #習慣	(6) [0.4%] Q:Aの習慣・習性
13								(17) [1.1%] R:習慣
14								(20) [1.3%] T:習慣
15								(21) [1.4%] S:Aの習慣・経験
16								(17) [1.1%] P:客観的事実2(Aの経験したこと)
17								(23) [1.5%] Q:経験
18								(2) [0.1%] S:Aの心の動き
19								(24) [1.6%] T:経験
20								
21								
22	(669) [44.6%] #:事実		(216) [14.4%] #:行動		(4) [0.3%] Q:過去			
23								
24								
25								
26								
27								
28								
29								
30								
31								
32								
33								
34								
35								
36								
37								
38								
39								
40								
41								
42								
43								
44								
45								
46								
47								
48								
49								
50								
51								
52								
53								
54								
55								
56								
57								
58								
59								
60								
61								
62								
63								
64								
65								
66								
67								
68								
69								
70								
71								
72								
73								
74								
75								
76								
77								
78								
79								
80								

第4章 雑談対話システムにおける言外の情報の有用性検証

本章では、第3章で明らかにした言外の情報の中で、雑談対話システムの構築に役立つ言外の情報が存在することを確認する。雑談対話において、本研究のように、言外の情報を扱うことを試みる研究は存在しないため、言外の情報が有用であることを示す必要がある。雑談対話システムにおいて、ユーザが「システムに理解されている」という印象を持つことは重要である。言外の情報をシステム発話として伝達できれば、ユーザは、システムがユーザ発話をどのように理解したかを把握でき、システムに理解されているという印象を持ちやすくなると考えられる。そこで、最も基本的な方法として、理解した言外の情報をそのままシステム発話としたとき、ユーザに及ぼす影響を調査し、有用な言外の情報が存在することを確認する。

第3章の調査の過程で、言外の情報のデータ、および、その類型が得られている。言外の情報の一部には、内容によっては、その伝達が好ましくない場合（例えば、相手を批判する内容や相手のプライバシーに関する内容など）も存在する。そのため、雑談対話システムにおいて、どのような言外の情報の伝達が好ましいか、好ましくないかを明らかにするため、言外の情報を伝達する発話の効果を分析する。具体的には、各類型の言外の情報を伝達するシステム発話を作成し、被験者実験を通じて、どのような類型の言外の情報であれば、その伝達が好ましいのか、好ましくないのかを調査する。

4.1 有用性検証の流れ

システムが言外の情報をユーザに伝達したときの効果を明らかにするため、言外の情報を用いて、評価対象となるシステム発話の作成、および、その評価を行う。以降、評価対象のシステム発話の作成方法、および、その評価方法について述べた後、評価結果とそれに基づく考察を述べる。

図4.1に、システム発話の作成と評価の流れを示す。まず、収集した言外の情報に基づいてユーザ発話と言外の情報のペアをランダムにサンプリングする。次に、サンプリングされたユーザ発話と言外の情報に基づいて複数種類のシステム発話を作成する。最後に、作成されたシステム発話を人手で評価し、その結果を分析することで、言外の情報を伝達する発話の効果を明らかにする。

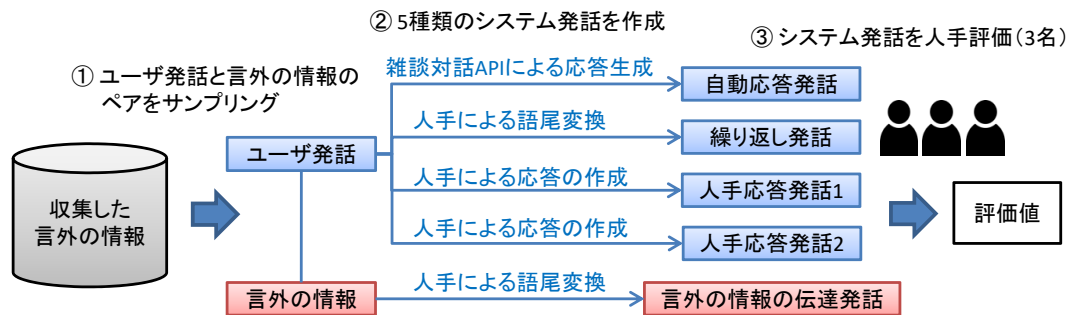


図 4.1: 言外の情報を伝達するシステム発話（言外の情報の伝達発話）の有用性検証

4.2 評価対象となるシステム発話の種類

表 4.1 に、評価対象のシステム発話の一覧を示す。言外の情報を伝達する発話（言外の情報の伝達発話と呼ぶ）を評価するために、比較対象の発話として他に 3 種類の発話（自動応答発話、繰り返し発話、人手応答発話）を用意した。計 4 種類の発話のうち、自動応答発話を除いた 3 種類の発話は全て人手で作成した。以下で各発話の作成方法について述べる。

言外の情報の伝達発話

この発話はユーザ発話の理解結果をそのまま伝達する発話である。発話の形式は、対話において自然だと考える「～ですね」を発話の末尾に付与した形式とする。この書き換え作業は人手で行う。このとき、元となるコーパス中の言外の情報に記述されている話者（A および B）は、発話としての自然さを損なわないように省略する。例えば、図 3.2 における P_2 の「B は車を運転する」という言外の情報を言外の情報の伝達発話に書き換える場合、「B は」を省略し、末尾を「ですね」と書き換えることで、「車を運転するんですね」という発話を作成する。

自動応答発話

この発話はユーザ発話を入力した際の、既存の雑談対話システムの応答発話である。ベースラインとして現状のシステムの応答レベルと比較することで、本手法でどの程度システム発話の評価値の改善が可能かを確認する。自動応答発話を作成するために、NTT docomo が提供している雑談対話 API¹ [56] を利用する。この API は、ユーザ発話を入力として、その中に含まれる焦点語（話題を表すキーワード）を抽出し、その焦点語に基づいて、大規模なテキストから構築された発話データベースに基づき応答を生成する。自動的

¹https://dev.smt.docomo.ne.jp/?p=docs.api.page&api_name=dialogue&p_name=api_reference. 2019 年 6 月をもって本 API の提供は終了し、katarai (<https://www.katar.ai>) と呼ばれる有料サービスに移行している。

に生成された発話であるため、一部不自然な発話も含まれる。例えば、図 3.2 における U_{13} の発話「気が向いたら、富士山に行ったりしますね。」に対しては、「富士山」が焦点語として抽出され、「富士山で働いていますか？」という発話が生成される。

繰り返し発話

この発話はユーザ発話に含まれる述語項構造に基づいて、ユーザ発話に含まれる内容を繰り返す発話である [27]。先行研究で述べられている通り、この発話は、ユーザ発話の表層のみを用いて生成が可能であり、かつ、ユーザに理解を示す効果がある。そこで、この発話を本研究におけるベースラインの発話とみなす。繰り返し発話と比較することで、言外の情報を伝達した際の効果を確認できる。発話は自動的に作成することも可能であるが、誤った自動解析結果に基づいて発話が作成されることを避けるため、人手で述語項構造を抽出し、それを書き換えることで発話を作成する。例えば、図 3.2 における U_{13} に対しては、主節から「述語：行く、二格：富士山」という述語項構造を抽出し、「富士山に行くんですね」というシステム発話を作成する。

人手応答発話

この発話はユーザ発話に対して人手で作成した応答発話であり、評価値の上限を調査するためのものである。発話作成者はユーザ発話中のキーワードを 1 つ選択し、そのキーワードを用いて適切な発話を記述する。適切な発話の定義は、ユーザ発話までの対話文脈を参照したとき、作成した人手応答発話に後続する対話ができるべく盛り上がるような発話とした。このような制限を設けない場合、話題を変える発話や単に相槌を示す発話が作成され、比較対象の発話との比較が難しくなることが考えられるため、発話作成時に必ずユーザ発話に含まれるキーワードを用いるようにする。例えば、図 3.2 における U_{13} に対しては、「富士山」がキーワードとなり、「富士山は素敵ですね」という発話を作成される。

4.3 システム発話の作成と評価の方法

評価対象のシステム発話を作成するために、まず、ユーザ発話と言外の情報のペアをサンプリングした。言外の情報を伝達したときの効果は、言外の情報の種類によって異なる可能性が考えられる。そのため、サンプリングの際には、言外の情報の種類による効果の違いを明らかにできるように、表 3.2 で示したレベル 3 の種類（信念 1、信念 2、願望 1、願望 2、属性、行動、環境、事実 1、事実 2 の 9 種類）の事例数が同数になるよう、種類ごとに言外の情報の一様ランダムサンプリングを行った。ペアとなるユーザ発話には、言外の情報の記述に利用された発話を利用した。すなわち、抽出された 500 事例に含まれる言外の情報は種類ごとに

表 4.1: 評価対象のシステム発話. システム発話の例は, ユーザ発話が「気が向いたら富士山に行ったりしますね」(図 3.2 の例における U_{13}) の場合, および, 言外の情報が「富士山は紅葉が有名」の場合を表す.

システム発話	概要	発話例
自動応答発話	既存の雑談対話システム (NTT docomo が提供している雑談 API を使用) に対して, ユーザ発話を入力したときの応答発話	富士山で働いていますか?
繰り返し発話	ユーザ発話の意味内容 (述語項構造) を抽出し, システム発話としてそのまま繰り返して確認する発話	富士山に行かれるんですね
人手応答発話	ユーザ発話中のキーワードを使って, 人手で最も自然かつ対話が盛り上がりと考えられる応答を記述した発話. 発話の作成者に依存して発話の品質が異なる可能性を考慮し, 2 名の作業者が発話 (人手応答発話 1, 人手応答発話 2) を作成する	富士山は本当に綺麗な場所ですよ
言外の情報の伝達発話	ユーザ発話に対する言外の情報を確認する発話	富士山は紅葉が有名なんですね

55 事例または 56 事例 (=500 事例 ÷ 9 種類) となる. このとき, 言外の情報とペアとなるユーザ発話には, 同一の発話が抽出される場合があり, 1 回抽出された発話は 201 発話, 2 回抽出された発話は 63 発話, 3 回抽出された発話は 18 発話. 4 回以上抽出された発話は 24 発話であった. そのため, 500 発話に含まれる異なりのユーザ発話数は 306 (= 201 + 63 + 18 + 24) 発話であった.

言外の情報の伝達発話および繰り返し発話は単純な書き換えで作成でき, 品質が作成者によって大きく変わらないと考えたため, 著者とは異なる 1 名の作業者が発話を作成した. 人手応答発話については, 発話の品質が発話作成者によって異なる可能性が考えられるため, 著者とは異なる 2 名の作業者 (作成者 1, 作成者 2) が発話を作成した. 2 名の作業者には, 過去に言語データ作成の経験を有する専門家を割り当てることで, 一定の方針に従って高品質な発話が作成されるようにした. その結果, ユーザ発話と言外の情報のペア 500 個のそれぞれに対して, 5 種類のシステム発話 (2 名の発話作成者が作成した 2 種類の手応答発話を含む発話) が作成された. したがって, 評価対象のシステム発話数は 2,500 発話 (500 ペア × 5 種類) となった.

各システム発話が対話においてどの程度好ましいかを評価するため, 各発話の自然さを 3 名の作業者が主観評価した. 評価者には, 各システム発話がユーザ発話

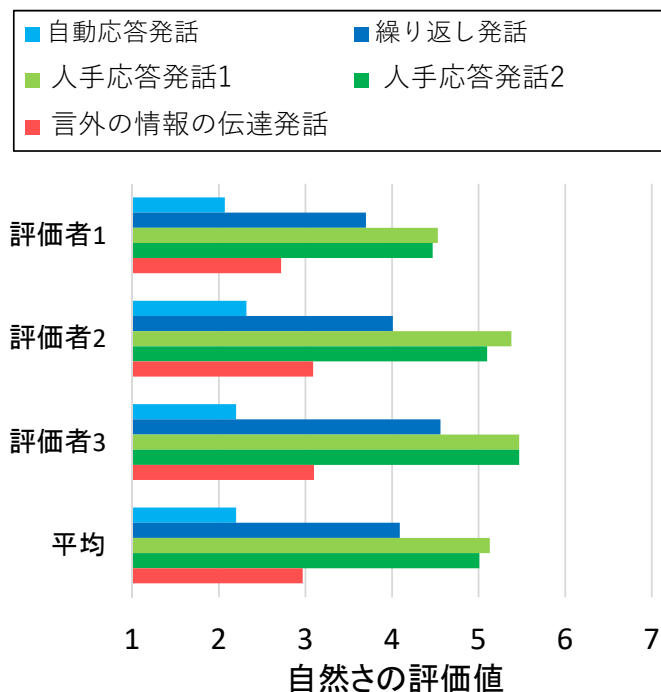


図 4.2: 3 名の評価者が付与した各システム発話の自然さの評価値

に対する応答としてどの程度自然かを評価するよう指示を与えた．評価値は 7 段階のリッカート尺度である（7 点：非常に自然～1 点：非常に不自然）．評価者はランダムな順序で並べられたシステム発話（直前のユーザ発話，および，それまでの対話文脈とともに提示される）を見て評価を行うようにした．今回の評価作業は，さまざまな言外の情報を伝達する発話や，人手で作成した適切な発話など，性質の異なる発話を一定の基準で評価する必要がある，一般的なシステム発話の評価作業と比較して複雑な評価であると考えられる．そのため，適切に評価が実施されるよう，評価者の 3 名には，著者とは異なる，テキスト評価の経験を有する専門家を設定した．各作業者は，全てのシステム発話に対して独立に評価を行った．評価作業は非専門家でも実施可能であると考えられるが，その妥当性は今後の検証が必要だと考える．

4.4 システム発話の主観評価結果

図 4.2 にシステム発話の評価結果を示す．図には各システム発話を 3 名の作業者が評価した結果，および，それらの平均が示されている．どの発話評価者による結果も，作成者 1 および作成者 2 が作成した人手応答発話が最も高い値になっており，自動応答発話が最も低い値となった．各システム発話に付与された評価値の順序は作業者間で一致しており，作成者 1 の人手応答発話，作成者 2 の人手応答発話，繰り返し発話，言外の情報の伝達発話，自動応答発話の順であった．作

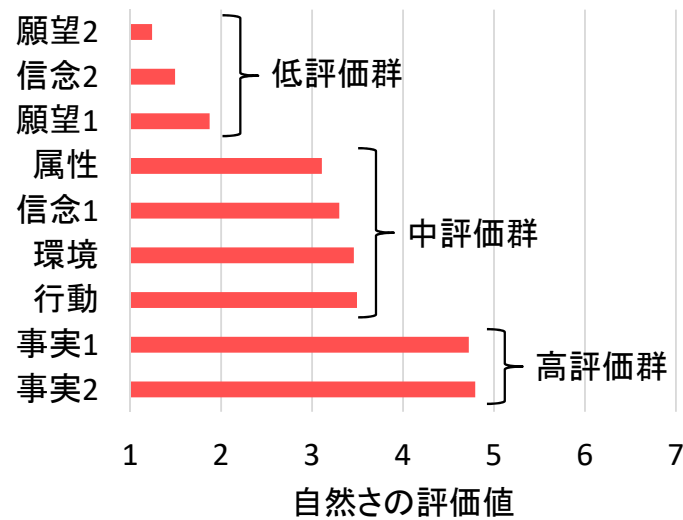


図 4.3: 言外の情報の種類ごとに言外の情報の伝達発話を分けたときの自然さの評価値

成者 1 の人手応答発話と作成者 2 の人手応答発話の間を除く任意のシステム発話の評価値の間で、統計的に有意な差が見られた (Steel-Dwass の多重比較において $p < .01$ [18])。作業者が各発話に対して付与した評価値の Spearman の順位相関係数は、全作業ペアの平均で 0.56 と比較的高い正の相関があり、各作業者が付与した評価値の傾向は類似していた。人手応答発話を評価値の上限とみなすと、繰り返し発話は述語項構造を繰り返す単純な発話であるが、高い評価値が付与されている。言外の情報の伝達発話は自動応答発話よりも高い評価値が付与されたものの、繰り返し発話よりも低い評価値であった。

4.5 言外の情報の類型に基づく主観評価値の分析

前節の結果から、言外の情報の伝達発話は繰り返し発話や人手応答発話と比べて自然さが低く評価されており、言外の情報をそのまま伝達する発話は概ね、ユーザにとって好ましくない結果となった。しかしながら、この結果は、言外の情報の類型によって異なる可能性が考えられる。そこで、言外の情報の類型による評価値の違いを明らかにするため、類型ごとに評価値を分けて調査を行った結果について述べる。

図 4.3 に言外の情報の伝達発話の類型ごと (表 3.2 のレベル 3) に評価値を分けた結果を示す。評価値は大きく 3 つのレンジ (1–2 点, 3–4 点, 4–5 点) に分かれた。以降、これらを低評価群、中評価群、高評価群と呼ぶ。低評価群に含まれる三つの種類の発話は、中評価群や高評価群の発話と有意な差が見られ (Steel-Dwass の多重比較において, $p < .01$)、低評価群内でも有意な差が見られた ($p < .05$)。中評価群に含まれる四つの種類の発話は、低評価群や高評価群の発話との間のみ

で有意な差が見られ ($p < .01$)、中評価群内では有意な差が見られなかった。高評価群に含まれる二つのタイプの発話も同様に、低評価群や中評価群の発話との間のみで有意な差が見られた ($p < .01$)。

表 4.2 および表 4.3 に、類型ごとに分けた言外の情報の伝達発話のうち、最も低い評価値が付与されたものと最も高い評価値が付与されたものを、言外の情報においてペアとなるユーザ発話とともに示す。以降、各タイプについて、表 4.2、および、表 4.3 を参照しながら、自然さの評価値に差が見られた理由について考察する。このとき、ポライトネス理論、および、情報のなわ張り理論との関連も述べながら考察する。

低評価群

低評価群の発話は、願望 2、信念 2、願望 1 の言外の情報を含んでおり、主にユーザの一時的な状態として、思考内容や感情に言及する発話である（例えば、「私に同意したいのですね」）。発話がポジティブな表現を含んでいたとしても、低い自然さが付与されている（例えば、「私のことが好きなのですね」）。低評価群に含まれる発話は、対話中のユーザの一時的な状態に言及しており、ユーザの状態を一方的に決め付けるような発話と言える。そのため、ネガティブフェイス（自由に行動したいという欲求）を侵害していると考えられる。また、システムの情報のなわ張りの外、かつ、ユーザのなわ張りの内の情報に直接形を用いて言及しているため、好ましくない印象を与えていると考えられる。特に、願望 2 や信念 2 のように、対話相手に対する思考内容に言及する発話は、評価値が低くなっており、ユーザにとってより好ましくないと考えられる。これは、願望 2 や信念 2 は願望 1 と比較して、より直近のユーザの思考に言及しており、それらの情報はシステムの情報のなわ張りからより遠く、ユーザの情報のなわ張りにより近い情報と位置付けられるからだと考えられる。なお、低評価群の中には「移動生活に憧れているのですね」というように、対話中の一時的な願望ではなく、普段からユーザが持つ願望が一部含まれており、それらは例外的に高い評価値が付与されている²。このことから、対話中の一時的な状態に言及する発話が好ましくない傾向があることがわかる。

中評価群

中評価群の発話は、属性、信念 1、環境、行動の言外の情報を含んでおり、主にユーザの興味、経験、プロフィールに言及する発話である（例えば、「かっこいい車が好きなんですね」）。この種類の発話は繰り返し発話と近い評価値が付与されていた。これらの発話はポジティブな内容を含んでおり、ユーザのポジティブフェイス（他者に承認されたいという欲求）を満たしているため高い評価値が付与されていると考えられる。低評価タイプ同様、システ

²例えば表 4.2 の願望 1 で最も高い自然さが付与された例のように、願望 1 には多様な願望が含まれており、例外的に高い値のものも存在する。

ムの情報のなわ張りの外、かつ、ユーザの情報のなわ張りの内の情報を含むが、ポジティブフェイスを満たしている場合、直接形でも好ましい印象を与えることが可能と考えられる。対話中の思考内容に言及する低評価群とは異なり、中評価群の発話はプロフィールなど、より客観的な内容に言及する発話であり、ネガティブフェイスの侵害は少ないと考えられる。加えて、ネガティブと取られうる表現を含む発話は自然さが低いと評価される傾向が見られた（例えば、「面倒くさがりですね」）。これは、ユーザのポジティブフェイスを侵害する発話である。すなわち、中評価群の場合、ポジティブな表現を含むものは高い評価値が、ネガティブな表現を含むものは低い評価値が付与される傾向があると言える。情報のなわ張り理論に照らし合わせると、ユーザに関するポジティブな情報は、ユーザの情報のなわ張りにおいても、より外に位置付けられ、ユーザに関するネガティブな情報は、より内に位置付けられる可能性があると考えられる。

高評価群

高評価群の発話は、事実1と事実2の言外の情報を含んでおり、ユーザに直接関係のない一般的な事実言及の発話である（例えば、「スポーツバイクは値段が高いですね」）。多くの発話は繰り返し発話よりも自然さが高く評価されており、人手応答発話と同等の評価値が付与されている。この結果から、一般的な事実に関する言外の情報であれば、それを伝達するシステム発話はユーザにとって好ましく、人手で最適な応答を作成した場合と同等の自然さになると考えられる。ポライトネス理論の観点から考えると、ユーザと直接関係のない一般的な事実を述べるだけであれば、ユーザのフェイスを侵害しないと考えられる。情報のなわ張り理論においては、高評価タイプに含まれる多くの発話は一般的な事実であり、システムにも関わりのある情報と考えられる。そのため、システムの情報のなわ張りの内となり、断定的な表現を用いて伝達しても好ましくなると考えられる。この結果から、高評価群に含まれるような一般的な事実に関する言外の情報をシステムが推定できれば、人手と同等レベルの自然さで理解内容の伝達ができる可能性がある。なお、中評価群とは異なり、ネガティブな表現を含んでいても、比較的高い評価値が付与されていたことも注目に値する。

高評価群に含まれる事実1および事実2の言外の情報は、表3.2の類型に示される通り全体の7%と他の類型と比較して多くない。また、事実1と事実2が記述された発話の割合を調査したところ、その割合は5%であった。他の類型と比較して、記述された数は少ない。しかしながら、これらの類型から可能となる発話は高い自然さとなっており、また、事実的な内容を推論する必要がある発話であり、システムの推論能力を示すことができる可能性もある。そのため、その割合以上に高い効果がある発話であると考えられ、雑談対話において有用であると考えられる。

表 4.2: ユーザ発話（矢印左側）と言外の情報の伝達発話（矢印右側）の例．図 4.3 と同様，言外の情報の類型は類型全体で評価値の低いものから順に並んでいる．言外の情報の類型ごとに，自然さの評価値が最低の事例が示されている．評価値には，作業者が付与した評価値の平均，および，各作業者が付与した値（丸括弧内）が示されている．

自然さ	類型	最も低い自然さが付与されたペア	評価値
低評価群	願望 2	困りますね → 私に共感してほしいのですね	1.0 (1, 1, 1)
	信念 2	おー！ → 私に好感を持ったのですね	1.0 (1, 1, 1)
	願望 1	映画を見にいつてきました → 話題を映画に変えたいのですね	1.0 (1, 1, 1)
中評価群	属性	車検されてしまい出れません → 面倒くさがりですね	1.0 (1, 1, 1)
	信念 1	キャビアを食べたことがないです → ちょっと恥ずかしいのですね	1.0 (1, 1, 1)
	環境	実はテレビは持ってないんですが → あなたの家には PC があるのですね	1.7 (2, 2, 1)
	行動	そうですね → どう伝えようか考えているのですね	1.3 (2, 1, 1)
高評価群	事実 1	(混雑で) 遠くから写真を撮りました → 周りに人が多かったのですね	4.0 (4, 4, 4)
	事実 2	パソコン触っているとあっという間です → パソコンは時間を奪うのですね	2.0 (1, 4, 1)

この考察の妥当性を確認するため，言外の情報の伝達発話に対して極性をラベルとして付与し，ラベルごとに自然さの評価値を確認した．極性のアノテーションは Stanford Sentiment Treebank の基準にならい，ポジティブまたはネガティブな表現を含むかどうかに基づき付与するようにした [78]．ラベルとしては，ポジティブ，ネガティブ，なし（極性なし）の 3 種類を用意した．作業は言語的アノテーションを専門とする作業員 2 名（著者とは異なる）が独立に行い，結果が異なるラベルは相談してどちらかに統一するようにした．

表 4.4 に，言外の情報の類型，および，極性で言外の情報の伝達発話を分けたときの自然さの評価値を示す．自然さの評価における低評価群では，極性のラベルによらず一様に低い自然さが付与されていることから，内容によらず好ましくない発話であることが確認できる．同様に高評価群でも，極性のラベルによらず好ましい発話であることが確認できる．中評価群では，属性と信念 1 について，ポジティブな内容を含む言外の情報の伝達発話（例えば，属性：「ワイルドだ」，信念 1：「カッコいい車が好き」）の自然さが，ネガティブまたは極性なし（例えば，属性：「面倒くさがりだ」，信念 1：「恥ずかしい」）と比べて評価値が高くなっている．このとき，属性および信念 1 について，ポジティブな内容を含む発話と，それ

表 4.3: ユーザ発話（矢印左側）と言外の情報の伝達発話（矢印右側）の例．図 4.3 と同様，言外の情報の類型は類型全体で評価値の低いものから順に並んでいる．言外の情報の類型ごとに，自然さの評価値が最高の事例が示されている．評価値には，作業者が付与した評価値の平均，および，各作業者が付与した値（丸括弧内）が示されている．

自然さ	類型	最も高い自然さが付与されたペア	評価値
低評価群	願望 2	(中国の話が) 私も気になってました → 私と中国の話をしたいのですね	2.0 (1, 2, 3)
	信念 2	そりゃそうですね → 私に同感なのですね	2.7 (3, 3, 2)
	願望 1	そんな暮らしができればいいなあ → 移動生活に憧れているのですね	4.7 (5, 2, 7)
中評価群	属性	スポーツバイクが趣味ですね → ワイルドですね	6.3 (5, 7, 7)
	信念 1	セダンやスポーツカーですが → かっこいい車が好きなのですね	6.3 (5, 7, 7)
	環境	クラブチームでバスケしてました → バスケチームに入っているのですね	6.0 (5, 6, 7)
	行動	この夏は蚊にたくさん刺されました → この夏よく外へ出かけたのですね	5.3 (3, 6, 7)
高評価群	事実 1	ホイール変えたので 5 万で → ホイールは値段が高いのですね	6.7 (7, 6, 7)
	事実 2	全部で 20 万ぐらいですね → スポーツバイクは値段が高いのですね	6.7 (7, 6, 7)

以外の発話について自然さを Steel-Dwass の多重比較にて検定したところ， $p < .01$ と有意な差が見られた．一方で，環境と行動（例えば，環境：「家に PC がある」，行動：「夏によく外に出かけた」）については，そのような違いは見られない．属性・信念 1 と，環境・行動の違いとして，前者はユーザ自身に関する事実（ユーザに尋ねないとわからないような事実）に言及しており，後者は外から判断可能なユーザに関する客観的事実に言及しているという点が挙げられる．情報のなわ張り理論においては，話し手に近い情報の定義に，話し手（システム）が体験によって得た情報が含まれる．したがって，環境・行動はシステムが体験によって得た情報として，属性・信念 1 と比較してより話し手に近い情報と位置付けられることから，極性による差が生じていないのではないかと考えられる．ここから，中評価群の結果については，自然さの評価値が低いことから言外の情報の伝達が好ましくないものの，例外的に，ユーザ自身（ユーザのプロフィールや考え）に対するポジティブな内容を含む言及については好まれる場合があると考えられる．

以上の結果に基づいて，自然さの観点から，言外の情報を伝達する際に好ましい，好ましくない情報をまとめると，ユーザの内面に言及する言外の情報はその伝達が好ましくなく，ユーザの内面以外では，ユーザ自身に言及する言外の情報

表 4.4: 言外の情報の類型、および、極性で言外の情報の伝達発話を分けたときの自然さの評価値。ポジ、なし、ネガは極性のラベルを表し、ポジはポジティブ、ネガはネガティブ、なしは「どちらでもない」に対応する。自然さの評価値は黄色が低い評価値、緑色が高い評価値を表す。極性の割合は言外の情報の類型ごとの極性の割合を表し、赤色が多いラベルを表す。

自然さ (群)	言外の情報 の類型	自然さの評価値			極性の割合		
		ポジ	なし	ネガ	ポジ	なし	ネガ
低評価	願望2	1.2	1.2	1.3	86%	10%	3%
	信念2	1.4	1.6	1.4	52%	34%	14%
	願望1	2.0	1.7	2.0	48%	38%	14%
中評価	属性	3.9	2.7	2.6	36%	43%	21%
	信念1	3.7	2.7	3.1	47%	20%	33%
	環境	3.5	3.4	3.6	22%	72%	5%
	行動	3.2	3.6	3.3	16%	71%	14%
高評価	事実1	5.3	4.5	5.0	10%	57%	33%
	事実2	5.3	4.6	4.4	38%	24%	38%

はポジティブな極性の内容のものであれば伝達が好ましくなる場合があり、ユーザに関する客観的事実やユーザと無関係な一般的事実（事実1および事実2）は極性によらずその伝達が好ましいということが明らかになった。今回得られた知見は、雑談対話システムの言外の情報がシステム発話として好ましいかを選別する基準となる有用なものであると考えている。

4.6 本章のまとめ

本章では、雑談対話システムにおいて、ユーザ発話の理解結果として、言外の情報を伝達する発話の効果进行分析した。具体的には、第3章で収集した言外の情報を用いて、言外の情報を伝達するシステム発話を作成、評価することで、雑談対話における言外の情報の有用性を検証した。その結果、自然さの観点から、ユーザの内面に言及する言外の情報はその伝達が好ましくなく、ユーザの内面以外では、ユーザ自身に言及する言外の情報はポジティブな極性の内容のものであれば伝達が好ましくなる場合があり、ユーザに関する客観的事実やユーザと無関係な一般的事実（事実1および事実2）は極性によらずその伝達が好ましいということが明らかになった。

今後の課題として、ユーザ発話からユーザにとって好ましい言外の情報を推定し、言外の情報を伝達してより円滑な対話を実現するシステムを構築したい。また、本研究では、言外の情報を基本的にそのまま伝達するシステム発話について分析を行なった。その結果、例えば、対話中のユーザの願望に関する発話は好ましくないという結果になった。しかし、例えば、ユーザの願望を満たすような発話（例えば、「ユーザは褒められたい」ということを理解した場合、ユーザを褒め

るような発話)を生成することで、好ましい発話が作成できる可能性がある。そのため、言外の情報の内容を考慮した発話生成についても取り組みたい。具体的には、雑談対話システムへの応用として、言外の情報をユーザ発話から推論したり、あらかじめ用意したシステム発話の候補が好ましいか否かを分類したりする手法が考えられる。例えば、自然言語文に対して、その前提や帰結を推論する手法 [67, 71] の応用が期待できる。加えて、自然言語理解のベンチマーク GLUE[93]での有効性が報告されている事前学習言語モデル [17, 65] は、常識的な推論の能力を備えているとされ [85]、本研究で扱う言外の情報の推論への応用が期待できる。また、システム発話候補を分類する手法として、その内容の極性が好ましさに影響を与えることから、Stanford Sentiment Treebank [78] に代表される極性判定の技術も重要である。

第5章 言外の情報に基づく議論対話システムの構築

本章では、言外の情報を考慮する雑談対話システムを構築する。上述した「言外の情報の有用性検証」で示したように、雑談対話システムにおいては、言外の情報の種類のうち、事実1および事実2の種類の伝達発話の自然さにおける評価値が最大となり、これらの種類の言外の情報の伝達発話が有用なことが明らかになった。そのため、言外の情報を考慮する雑談対話システムは、事実に基づく対話、かつ、今後人間との共同作業につながる実用性を伴う対話を扱うことが適切だと考えられる。そこで本研究では、オープンドメインの議論対話を扱う。議論対話は、自分の考えを述べたり、相手の考えに対して支持や反論の理由を述べたりする対話であり、この支持や反論は事実に関する言外の情報と考えられる。また、議論対話は意思決定や合意形成につながり、実用性のある対話だと考えられる。

本章の内容は、第4章までの内容とは間接的に関連するものであり、扱う対話内容や、評価の方法が異なる。これは、第3章と第4が一般的な雑談対話を対象にした研究であるのに対し、本章はこれらの取り組みから得られた知見を間接的に利用して雑談対話の内容を議論に限定して対話システムの構築に取り組んだためである。

5.1 タスク設定

図5.1に、本研究で取り組む議論対話システムのタスク設定を示す。入力として、特定のトピック（この例の場合は「料理」）に関する意見を述べるユーザ発話（以降、主張と呼ぶ）を受け付ける。この主張は、トピックに対する賛成意見や反対意見など、支持や反論が可能な議論につながる発話とする。システムは主張に対して、その内容を支持する理由や、その内容に反論する理由を述べる発話（以降、論拠と呼ぶ）を出力する。論拠は主張に対する支持や反論を述べるだけでなく、その理由を論理的かつ具体的に述べる発話とする。本研究では、この、主張となる発話に対し論拠となる発話を出力する問題に取り組む。ユーザが興味を持つトピックはさまざまであるため、任意のトピック（オープンドメイン）を対象とする。任意の主張に対して論拠を生成することができれば、ユーザが興味を持つさまざまなトピックに関する議論が可能となる。支持の論拠は「言外の情報の

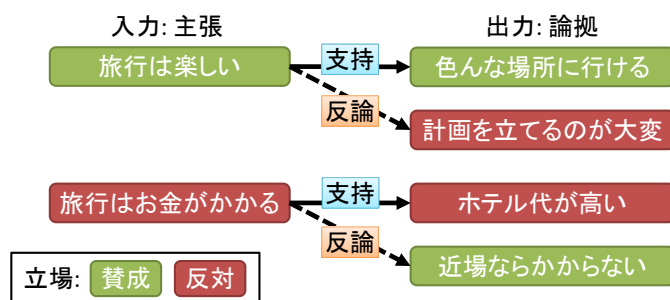


図 5.1: 議論対話システムのタスク設定（入力・出力の例）

収集と類型化」で示したように、入力に対する事実 1 または事実 2 の類型の言外の情報として考えられ、これらの類型は「言外の情報の有用性検証」で示したように、システム発話として有用である。また、システム側の意図に応じて、支持ではなく反論を返すべき状況（例えば、ユーザの主張がシステムの意図と反する場合）も存在すると考えられるため、反論にも取り組む。

5.2 主張・論拠ペアデータの人手収集

本節では、議論対話システムにおける発話生成モデルを学習、および、評価するために人手で収集した主張・論拠ペアデータについて述べる。さまざまなトピックに関する主張、および、それに対する支持の論拠と反論の論拠を収集するために、クラウドソーシングを用いて作業者に主張や論拠を記述してもらうことでデータを収集した。以降、その収集手順、および、収集されたデータの規模と分析結果について述べる。

5.2.1 収集手順

図 5.2 に、主張・論拠ペアデータの人手収集手法を示す。各作業者は、＜トピック、主張、支持の論拠、反論の論拠＞の 4 つ組を 3 つずつ作成する。具体的には、まず、議論のトピックを名詞で 3 つ考えてもらい、その後、トピックに関して意見を述べる発話を 1 発話、および、その意見に対する支持の論拠と反論の論拠を 1 発話ずつ記述してもらうようにした。以降、詳細な手順について説明する。

Step 1. 議論のトピックの列挙

この手順では、作業者は自身の好きなもの、嫌いなもの、気になっているもの、またそれ以外思いついたものを名詞として 3 つ記述する。このとき、可能であれば固有名詞（例えば、特定の人名や地名など）をなるべく避けるように作業者に指示した。これは、一般名詞を優先してデータを収集することで、幅広い主張に対して支持や反論のデータを収集するためである。

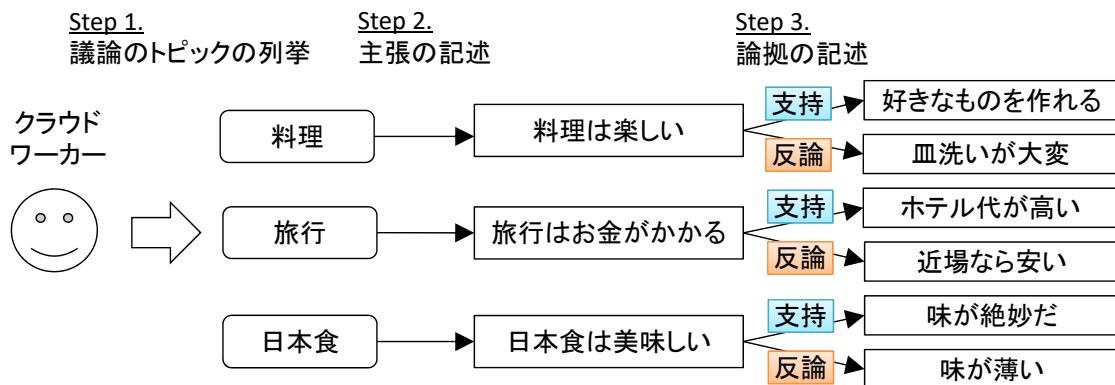


図 5.2: 主張・論拠ペアデータの手収集方法

Step 2. 主張の記述

この手順では、作業者は Step 1 で列挙したそれぞれの名詞を用いて、主張を記述する。主張は話し相手に対して、議論を呼ぶような発話になるようにし、意見を端的に述べる 1 文で記述するように支持した。記述される主張は主語と述語を必ず含むようにし、極端に短かい発話が作成されることを防ぐため、日本語で 5 文字以上になるよう制限を与えた。

Step 3. 論拠の記述

この手順では、作業者は主張に対する支持を表す発話と反論を表す発話を論拠として記述する。記述される論拠は必ず具体的な理由に基づいて主張に対する支持や反論を述べる発話であり、主張と異なる内容語を必ず含めるよう教示を与えた。支持や反論の論拠を考えるために、参考として作業者には支持であれば「(主張)。そうだよね、(論拠)。」、反論であれば、「(主張)。そうは思わない、(論拠)。」というテンプレートに当てはまる形で発話を考えるよう指示した。論拠についても、主張同様 5 文字以上の日本語になるよう制限を与えた。

付録の 5.6.1 節における図 5.6、および、図 5.7 に、クラウドソーシングの作業者に提示する教示、議論のトピックの列挙用記述フォーム、主張と支持および反論の論拠の記述用フォームを示す。クラウドソーシングでデータを収集する際は、収集するデータの品質を高めるため、タスクの内容を容易に理解することができ、個々の作業の難易度をなるべく簡便にするためにステップに分割することが重要となる。クラウドソーシングに参加する作業者は、図 5.6 で示された作業概要を表す教示を読み、議論のトピックを列挙した後、図 5.7 で主張と論拠を記述する。この教示および記述フォームは、複数回のトライアル作業を経て問題点の修正を行い、最終版を作成した。その結果、エラー回答（「あああ」などの無効な回答を含む作業結果）の割合を数%に抑制することができた。なお、クラウドソーシングを利用する作業では、今回の作業のように高度な知的作業を要する作業では一般的に作

表 5.1: 収集された主張・論拠ペアデータの統計情報

	支持			反論		
	主張	論拠	全体	主張	論拠	全体
発話数	45,000	45,000	90,000	45,000	45,000	90,000
トピック数	6,313	NA	NA	6,313	NA	NA
平均単語数	7.16	8.64	7.90	7.16	8.55	7.85
平均文字数	12.8	14.2	13.5	12.8	13.8	13.3
語彙数	10,892	12,666	16,540	10,892	12,119	16,217

業者に事前にテストを実施することで品質を高めるのが一般的であるが、本研究ではなるべく収集データの規模を大きくするよう、事前テストを設定しなかった。

5.2.2 収集された主張・論拠ペアデータ

表 5.1 に収集された主張・論拠ペアデータの統計情報を示す。“全体”の列に支持と反論として収集された発話の総数が示されている。15,000 人の作業者が Step 1 から Step 3 の作業を行うことで、支持と反論でそれぞれ 45,000 ペアの主張と論拠が収集された。“主張”と“論拠”の列に示されている通り、支持と反論の両方で主張よりも論拠の方が単語数が多くなっており、より詳細な情報が理由として含まれていると考えられる。

5.2.3 収集された主張・論拠ペアデータの分析

収集された主張と論拠において、どのような発話が多く見られるのかを分析した。分析の観点として、発話を分類する基本的な体系である対話行為を利用した。また、議論における典型的な支持や反論の方法論をまとめた理論である Walton の議論スキーム [92] との関連についても述べる。対話行為の定義として、日本語の雑談対話を対象に設計された Meguro らの体系 [49] を利用した。この体系は発話をおよそ 30 種類の対話行為に分類するものであり、大規模な雑談対話データを利用して設計されている汎用的な体系であることから、本研究の分析に利用する上で適切である。

表 5.2 に、主張・論拠ペアデータにおける頻出の対話行為を示す。これらの対話行為は、Meguro らの体系を用いて学習されたロジスティック回帰を用いた分類器を用いて推定されたものである。分類器の性能は正解率が 49%であり、人手による対話行為のアノテーション一致率が 59%であることから、十分な性能を持つ分類器である。まず“主張”に着目すると、支持と反論の両方において、“自己開示: 評価+”，すなわち、議論のトピックに対する肯定的な意見が頻出していることがわかる。支持の論拠に着目すると、主張と同じ対話行為である“自己開示: 評価+”

表 5.2: 主張・論拠ペアデータにおける頻出の対話行為. Meguro らの体系 [49] において, + はポジティブな極性の評価, - はネガティブな極性の評価を表す.

		主張	論拠	割合
支持	1	自己開示: 評価 +	自己開示: 評価 +	33%
	2	自己開示: 評価 +	情報提供	14%
	3	自己開示: 評価 +	自己開示: 事実	11%
	4	自己開示: 事実	自己開示: 評価 +	3%
	5	情報提供	情報提供	2%
反論	1	自己開示: 評価 +	自己開示: 評価 -	23%
	2	自己開示: 評価 +	自己開示: 評価 +	19%
	3	自己開示: 評価 +	自己開示: 事実	10%
	4	自己開示: 評価 +	情報提供	3%
	5	自己開示: 評価 -	自己開示: 評価 +	2%

が最も多く出現している. これらの事例を分析した結果, Walton の議論スキームにおける, Argument from Composition (具体例を用いる論拠), Argument from Division (抽象概念を用いる論拠), Argument from Consequences (帰結の善し悪しを用いる論拠) などが見られた. 一方, 反論の論拠に着目すると, 支持とは異なり, “自己開示: 評価 -” と “自己開示: 評価 +” の 2 種類が多く出現していることがわかる. これらの事例を分析した結果, “自己開示: 評価 +” では, 支持と同じ議論スキームである Argument from Composition, Argument from Division, Argument from Consequences や, Argument from Exception (例外を用いる論拠) が多く見られた. 一方で, “自己開示: 評価 +” では, Argument from Opposition (比較を用いる論拠) が多く見られた. この結果から, 収集された主張・論拠ペアデータにはさまざまな種類の支持や反論が含まれていることがわかる. また, 支持や反論を生成する上で, 主張や論拠の肯定や否定を適切に理解することが重要であると考えられる.

5.3 ガイドと主張・論拠ペアに基づく発話生成モデル

本研究では, さまざまなトピックに言及する主張に対して適切に支持や反論を生成するために, 主張や論拠の立場を考慮するガイドを導入し, さらに, さまざまなトピックを含む主張・論拠ペアを自動的に獲得し学習に利用する手法を提案する.

ガイド付き主張・論拠ペアデータは, 二種類の方法 (自動収集と人手収集) で作成される. 自動収集では, 大規模テキストから自動的に主張・論拠ペアデータを獲得する. 品質は完全ではないものの, さまざまなトピックに関するペアを獲得することができる. もう一方の人手収集では, 先に述べたクラウドソーシング

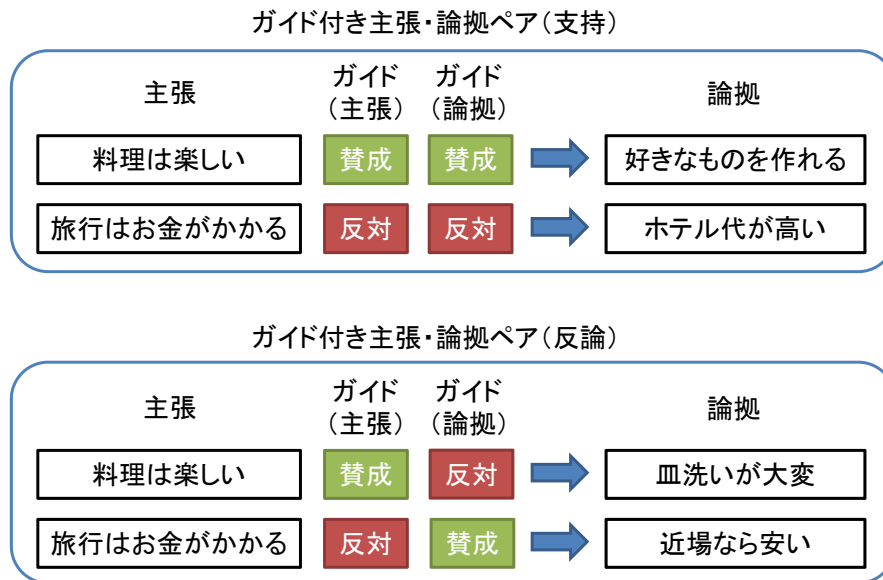


図 5.3: 発話生成モデルの学習に用いるガイド付き主張・論拠ペアデータの形式

を用いて人手で記述した主張・論拠ペアデータを用いる。規模は限られるものの、高品質なペアを収集することができる。これらのデータを用いて、自動収集したデータで第一段 fine-tuning を行った後、人手収集したデータで第二段 fine-tuning (パラメータの精緻化に対応する学習) を行う。このように、複数回 fine-tuning を行う手法は発話生成モデルを学習する効果的な手法として知られている [62]。

5.3.1 主張と論拠の立場を考慮するガイド

図 5.3 に、発話生成モデルの学習に用いるガイド付き主張・論拠ペアデータの形式を示す。図中には支持と反論のガイド付き主張・論拠ペアの例が示されており、これらの形式は支持と反論で共通である。本研究では、支持または反論のペアを正確に発話生成モデルに学習させるために、ガイドを導入する。入力となる主張には、トピックに賛成する立場の発話（例.「料理は楽しい」）だけでなく、トピックに反対する立場の発話（例.「旅行はお金がかかる」）も含まれる。そのため、出力される論拠は、図に示すように、これらの立場を考慮し、支持であれば主張と同じ立場、反論であれば主張と異なる立場でなければならない。本研究では、この立場の情報をトークン（発話生成モデルが扱う入力および出力に含まれるシンボルの最小単位）として入力に追加することで、正確に支持や反論を生成する。ガイドをトークンとして追加する手法は簡便であるものの、出力を操作する手法として有効であることが報告されている [55, 65]。

ガイドとして、主張と論拠のそれぞれについて、賛成、反対、なしの 3 種類をトークンとして用意する。以降説明する学習用の主張・論拠ペアデータの作成で

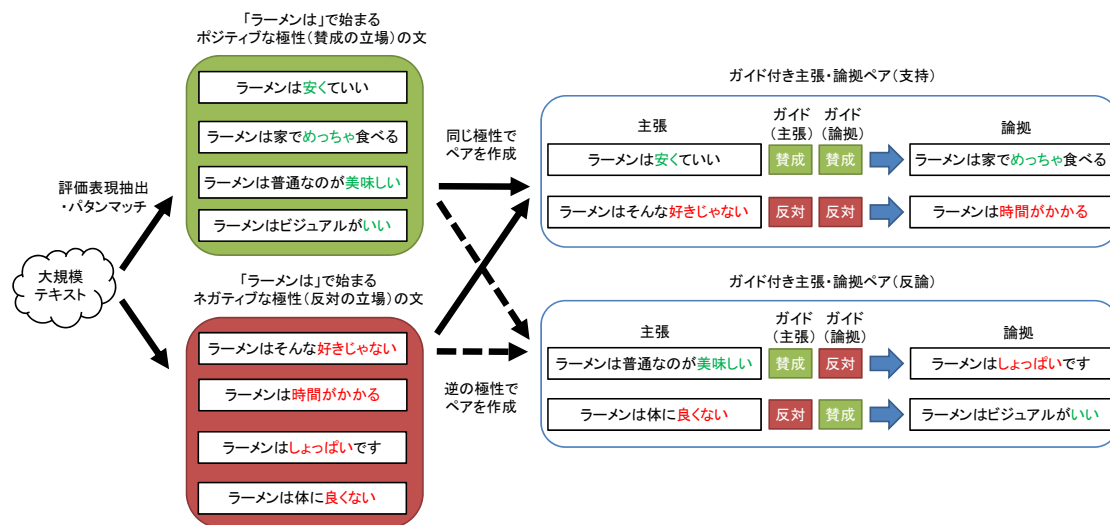


図 5.4: ガイド付き主張・論拠ペアデータの自動獲得

は、ポジティブな評価表現を含むものを賛成とし、ネガティブな評価表現を含むものを反対とする。そのどちらも含まない場合や両方を含む場合はなしとする。

5.3.2 ガイド付き主張・論拠ペアの自動獲得

図 5.4 に、ガイド付き主張・論拠ペアデータの自動収集手法を示す。本手法では、支持または反論の関係に近いと考えられる文のペアを大規模に獲得する。ペアを獲得する一般的な手法として、2 文の関係を分類器で推定し、支持や反論の関係にある文を獲得する手法や、支持または反論を表す接続詞（「したがって」や「しかし」）で繋がる 2 文を獲得する手法が考えられるが、精度やカバレッジが十分とは限らない。そこで、本研究では、特定のトピックに言及する文であり、かつ、極性を持つ文をランダムにペアにすることで、ペアデータを作成する。この手法を用いることで、さまざまなトピックに関する支持や反論らしい表現をペアデータとして獲得できると考えられる。

具体的な方法として、まず、大規模テキストに対して評価表現抽出を行い、極性を持つ文を抽出する。このとき、どのトピックに言及する文かを明らかにするため「X (名詞) + は」で始まる文のみを抽出する。次に、極性をトピックに関する立場（賛成か反対か）とみなし、支持であれば同じ極性、反論であれば異なる極性の文をペアとして組み合わせることで、文ペアを作成する。以降、各手順について詳細に説明する。

まず、特定のトピックに対する主張や論拠と考えられる意見となる文を抽出する。このとき、抽出対象の文には、要件として、意見を表す肯定や否定の表現が含まれており、さらに、どのトピックに対する言及であるかが明確である必要がある。前者の要件を満たすために、評価表現抽出では、大規模テキスト中の各文

に対して評価表現抽出器を適用し、評価表現を含む文を抽出する。評価表現抽出器として、誤字などを含む Web テキストにも頑健に抽出が可能と考えられる辞書ベースの手法 [38] を利用した。評価表現として、ポジティブな極性のもの、および、ネガティブな極性のものが正確に判定可能なもののみを抽出し、文意に応じてどちらとも考えられるもの（例えば、「高い」など。Kobayashi らの手法では、どちらの極性も取りうる表現として抽出される）は抽出しないようにした。後者の要件を満たすために、パターンマッチを用いて、特定のトピックに対して言及していると考えられる文のみにフィルタする。このためのパターンとして、トピックを表す名詞の直後に、提題を表す助詞の「は」が続き、その後ろにポジティブまたはネガティブな極性を持つ評価表現が含まれる文を設定した。このパターンはシンプルであるが、トピックに関して特定の分類器などを利用しておらず抽出結果が分類器の学習データに依存しないことから、さまざまなトピックに対して言及するさまざまな表現を含む文を頑健に抽出できると考えられる。これらの評価表現抽出とパターンマッチの結果、特定のトピック（例。ラーメン）に関するポジティブな極性（賛成の立場）の文、および、ネガティブな極性（反対の立場）の文を抽出することができる。

次に、抽出された文を用いて、支持と反論のガイド付き主張・論拠ペアを作成する。支持のペアを作成する際には、同じ極性の文をランダムに組み合わせることで、主張の文と論拠の文が支持らしい関係になるようにする。一方、反論のペアを作成する際には、異なる極性の文をランダムに組み合わせることで、主張の文と論拠の文が反論らしい関係になるようにする。このとき、作成される主張・論拠ペアに含まれる文には、主張と論拠を合わせて、同じ文が複数回出現しないようにする。これは、予備的にランダムに文を選択しペアを作成した場合、発話生成モデルの学習が適切に収束しなかったためである。作成された主張・論拠ペアについて、主張と論拠の極性を用いて、ガイドを導入する。このようにした作成されたペアをガイド付き主張・論拠ペアデータとみなす。

5.3.3 ガイド付き主張・論拠ペアの人手獲得

図 5.5 に、ガイド付き主張・論拠ペアデータの人手収集手法を示す。図中のトピック、主張、論拠は、図 5.2 で収集されたものを表している。収集された主張と論拠に対して、それぞれ自動獲得同様に評価表現抽出を行うことで、それぞれのガイドをトークンとして主張に不可することで学習データを作成する。このとき、主張や論拠に評価表現が含まれていない場合は、立場としてなしのトークンを追加するようにした。

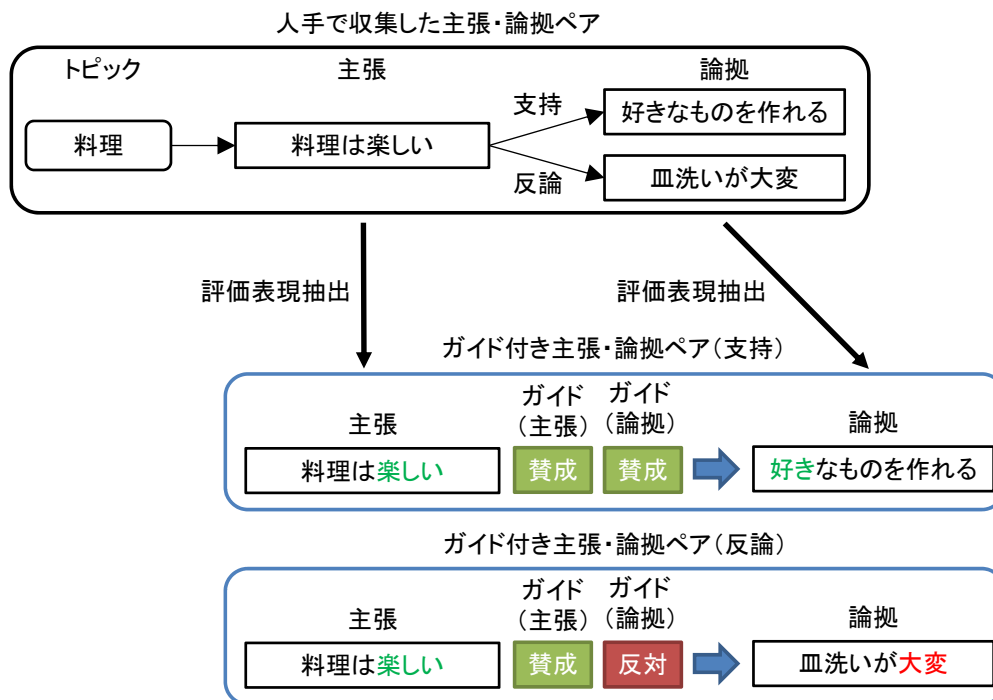


図 5.5: ガイド付き主張・論拠ペアデータの人手獲得

5.4 評価実験

提案手法の有効性を確認するため、評価実験を行った。具体的には、主張・論拠ペアデータを用いて作成した2種類の主張を入力として用意し、ベースラインと提案手法を導入した発話生成モデルを学習した後、各手法から生成された論拠を評価した。

5.4.1 評価対象の発話生成手法

表 5.3 に、評価対象の発話生成手法で利用する素性の一覧を示す。ベースラインとして利用する Baseline に加えて、+Pol, +Pre, +Swi の3種類の素性を比較のために用意した。

Baseline

Baseline は、ベースラインとして利用する発話生成手法であり、BERT[17]を入力発話の encoding, および、出力発話の decoding に利用する手法である[13]。BERT はそのままでは発話生成に利用することはできないが、Correia らは encoding 用の BERT と decoding 用の BERT のパラメータを共有し、さらに、decoding 用の BERT を seq2seq 同等の発話生成モデルとして fine-tuning することで発話生成に適用可能にしている。利用される BERT は大規模なテキストデータから学習されたものであるため、本論文で提案する各

表 5.3: 発話生成手法として利用する素性

素性	概要	必要性
Baseline	ベースラインとなる BERT ベースの手法 [13]. 人手獲得した主張・論拠ペアを用いて fine-tuning を行う.	必須
Pre	自動獲得した主張・論拠ペアを用いて事前に fine-tuning を行う手法 (提案手法)	オプション
Pol	主張と論拠の立場をガイドとして利用する手法 (提案手法)	オプション
Swi	支持と反論の学習データを併合し, 支持と反論のトークンに基づき同時に学習する手法	オプション

素性が BERT のような事前学習言語モデルで捉えられていない情報をモデル化するために有効であるかを確認することができる.

+Pre

+Pre は, 学習データとして, クラウドソーシングで収集したデータだけでなく, 自動獲得したデータを用いて事前学習する手法である. 学習の際は, 自動獲得したデータを用いて発話生成モデルを fine-tuning した後, クラウドソーシングで収集したデータを用いて発話生成モデルをさらに fine-tuning する.

+Pol

+Pol は, 立場のガイドを利用する手法である. 学習の際は, 学習データに含まれる出力となる論拠のガイドをそのまま利用する. 一方テストの際は, 真の出力となる論拠の情報は利用できないため, 入力となる主張の情報を元に論拠のガイドを推定して生成に利用する. 具体的には, 入力となる主張に対して評価表現抽出 [38] を行い, ポジティブまたはネガティブな極性がどちらかのみ 1 つ以上推定された場合は, 支持の場合は主張と同じ極性のガイドを論拠のガイドとし, 反論の場合は主張と異なる極性のガイドを論拠のガイドとする. 主張に対して極性が抽出できなかったり, ポジティブな極性とネガティブな極性の両方が抽出された場合は, 主張と論拠の両方のガイドになしを設定する.

+Swi

+Swi は, 支持と反論の学習データを併合し, 支持と反論の生成を同時に学習する手法である. 支持と反論の出力については, ガイドと同様に特殊なトークンを支持と反論の各々について用意し, 入力の最後に不可することで制御する. 支持と反論を個別にした場合と比較して, 学習データの規模がおおよそ

2 倍になるため、よりさまざまなトピックに関する手法に頑健に支持や反論を生成できる可能性がある。

5.4.2 モデルの学習と評価の方法

クラウドソーシングを用いて収集した主張・論拠ペアデータは、発話生成モデルの学習、および、発話生成モデルを評価するための入力となる主張、および、リファレンスの出力となる論拠としてして利用される。これを Crowdsourcing と呼ぶ。表 5.2 に示したように、クラウドソーシングを用いて収集した主張・論拠ペアデータは収集された主張の立場が賛成の事例が多いため、このデータだけを用いて評価した場合、反対の立場の主張に対して適切に支持や反論の論拠を生成できるかが正しく評価できるとは限らない。そこで、クラウドソーシングで収集した主張・論拠ペアデータの入力だけでなく、そのデータを用いて、賛成と反対の立場をバランスさせて擬似的に主張を作成した入力を Artificial として用意した。

Crowdsourcing

クラウドソーシングを用いて収集した主張・論拠ペアデータを利用する入力である。発話生成モデルの学習、および、後述する自動評価、および、人手評価の両方に利用する。クラウドソーシング収集された発話には、一部エラーが含まれていたため、適切なデータのみをフィルタし実験に利用した。具体的には、主張には議論のトピックを表す単語が文字列として含まれており、また、その他に述語となる内容語が 1 語以上含まれており、かつ、論拠にも内容語が 1 語以上含まれているもののみを適切な主張・論拠ペアとしてフィルタした。このフィルタは、支持と反論の両方で行ない、どちらか一方でも条件を満たさなかった場合は不適切とした。その結果、ペアの数は収集した 45,000 ペアのうち 20,767 ペア (46%) となった。発話生成モデル学習のため、このデータを 8:1:1 に分割し、学習用データ、開発用データ、テスト用データとした。テスト用データのうち、人手評価のために 272 発話をランダムに選択した。このとき、議論の発話生成に取り組んだ先行研究 [11, 59] にならって、未知のトピックに対する論拠の生成を評価するよう、トピックが重複しないように学習データを分割した。

Artificial クラウドソーシングを用いて収集した主張・論拠ペアデータから、賛成と反対の立場をバランスさせた主張を擬似的に作成した入力である。発話生成モデルの学習、および、後述する自動評価には利用せず、人手評価のみに利用する。Crowdsourcing に含まれるテストデータのトピックを用いて、賛成の立場の主張として「(トピック) が好き」、反対の立場の主張として「(トピック) が嫌い」というテンプレートを用意し、同一のトピックを入れることで 2 種類の入力を作成した。Crowdsourcing 同様 272 発話を人手評価で利

用したため、その半分の 131 トピックを用いて入力を作成された

Pre を学習するために、独自にクロールした 300GB (2 億文書) の大規模テキストコーパス (2016 年 1 月 1 日から 2016 年 12 月 31 日の 1 年間に記述されたもの) を利用した。その結果、支持で 740,842 ペアの反論で 839,054 ペアの主張・論拠ペアが獲得され、含まれるトピック数はそれぞれ 37,174 トピック、および、30,023 トピックとなり、クラウドソーシングで収集したトピック数 (表 5.1 に記載されている数値) と比較して、支持と反論でそれぞれ 588%, 475% となった。この結果から、提案手法を利用することでさまざまなトピックを含む学習データを獲得できることが確認できる。

Baseline として利用する BERT のモデルとして、日本語 Wikipedia で学習された事前学習モデル¹を利用した。トークナイザとして、同じく日本語 Wikipedia で学習された sentencepiece[39] を利用した。

表 5.4 に、評価対象の発話生成手法と学習方法の対応関係を示す。表に示される各発話生成モデルを学習するために、fine-tuning として学習データごとに 20 epoch の fine-tuning を行い、開発用データにおける perplexity を用いて学習を止めた。学習データに含まれるトークンの最大長は 32 トークンとし、学習率とドロップアウト率をそれぞれ 5×10^{-5} 、0.1 とした。

表 5.4: 評価対象の発話生成手法と学習方法の対応関係

比較手法	第一段 fine-tuning (自動収集した主張・論拠ペアを利用)	第二段 fine-tuning (人手収集した主張・論拠ペアを利用)	支持と反論 の同時学習
Baseline	なし	あり (ガイドなし)	なし
+Pre	あり (ガイドなし)	あり (ガイドなし)	なし
+Pol	なし	あり (ガイドあり)	なし
+Swi	なし	あり (ガイドなし)	あり
+Pre+Swi	あり (ガイドなし)	あり (ガイドなし)	あり
+Pol+Swi	なし	あり (ガイドあり)	あり
+Pre+Pol	あり (ガイドあり)	あり (ガイドあり)	なし
+Pre+Pol+Swi	あり (ガイドあり)	あり (ガイドあり)	あり

5.4.3 自動評価

対話システムにおける応答発話の自動評価尺度として、代表的なものに BLEU[58] がある。BLEU とは、機械翻訳における翻訳文を評価する尺度として提案された

¹<https://github.com/yoheikikuta/bert-japanese>

表 5.5: 各発話生成手法で Crowdsourcing における支持および反論の論拠を生成したときの自動評価結果 (perplexity). +Pre+Pol は提案手法である. 太字は最も良い評価値を表す.

発話生成手法	Crowdsourcing	
	支持	反論
Baseline	8.058	7.336
+Pre	5.741	5.383
+Pol	8.245	7.773
+Swi	7.150	6.579
+Pre+Swi	5.596	5.231
+Pol+Swi	7.301	7.112
+Pre+Pol(提案手法)	6.003	6.199
+Pre+Pol+Swi	5.874	5.861

ものであり、システムが出力した翻訳文の単語列が、正解となる翻訳文の単語列と一致している程度を評価値として利用する尺度である。機械翻訳の分野では、システムが出力した翻訳文に対して付与された BLEU における評価値と、人手による評価値の相関が高いことが報告されている [57]。一方で、対話システムにおいては、BLEU と人手による評価値との相関が高くないことが報告されている [46, 47]。そのため、近年の対話システムにおける発話生成の研究では、自動評価尺度として perplexity を利用することが多い [1, 80]。perplexity とは、テストデータにおける正解をどの程度曖昧性なく生成できるかを表す尺度であり、発話生成モデルが出力する生成発話におけるトークンの確率分布（次にどの単語が最も来やすいか）に基づいて計算される基本的な尺度である。perplexity は、その値が低いほど発話生成モデルがテストデータにおける正解の発話を曖昧性なく出力できていることを表すため、より低い値が良い結果を表す。そのため、本研究でも perplexity を用いて評価する。

表 5.5 に、発話生成モデルで Crowdsourcing における支持および反論の論拠を生成したときの自動評価結果 (perplexity) を示す。支持、反論の双方において、提案手法である +Pre+Pol が Baseline よりも低い perplexity を示しており、提案手法が論拠の生成に有効であることが確認できる。

最も低い perplexity を示しているのは +Pre+Swi である。提案手法である +Pre+Pol や、全ての素性を利用した +Pre+Pol+Swi よりも低い値を示しており、この結果は、Pol の素性が perplexity の観点では必ずしも有用でないことを表す。同様に、この結果は、Baseline と素性を 1 種類だけ利用するモデル (+Pre, +Pol, +Swi) を比較したとき、+Pol のみ、perplexity が低くなっていないことから確認できる。理由として、+Pol が使用する主張のガイドと論拠のガイドにおいて、論拠のガイドが主張のガイドに基づいて自動的に推定されたものであり、必ずしも Crowdsourcing

表 5.6: Crowdsourcing のリファレンスの情報に基づいてガイドを与えた場合の、開発データにおける+Pol を用いる各発話生成手法の自動評価結果（perplexity）

発話生成手法	Crowdsourcing	
	支持	反論
Baseline	8.644	7.833
+Pol	8.377	7.367
+Pol+Swi	7.230	6.693
+Pre+Pol(提案手法)	5.749	5.246
+Pre+Pol+Swi	5.514	5.093

における正解の論拠のガイドと同じとは限らないことに起因すると考えられる。+Pol で使用されているガイドは、指定した立場（賛成または反対）として適切な論拠の生成を促す情報である。そのため、リファレンスの立場と同じ論拠のガイドを指定した場合、perplexity を低下させることができるが、リファレンスの立場と異なる論拠のガイドを指定した場合、リファレンスを生成する観点ではノイズとなり、perplexity を向上させる。そのため、人手で判断したとき適切な支持や反論が生成できていたとしても、リファレンスが生成されにくくなった場合、perplexity としては悪化する。そのため、出力となる論拠のガイドを明示的に与えている+Pol は perplexity が高くなっていると考えられる。特に、perplexity の悪化は支持と比較して反論の方が顕著である。このことは、表 5.2 で示したように、支持と反論における出力となる論拠の立場を比較すると、反論の方が出力の立場の種類が多様である（支持は主張と論拠の立場が同じ割合が 33%であるのに対し、反論は 23%と割合が低くなっている）ことから示唆されることが考えられる。

Baseline と+Swi を比較すると、支持と反論の両方で+Swi の方が perplexity が低くなっており、学習データが増えることでより適切な支持や反論が生成されやすくなることが確認できる。また、+Swi と+Pre の比較、および、+Pol+Swi と+Pre+Pol の比較から、Swi と比較して Pre を利用した方が perplexity が低下していることがわかる。この結果は、学習データを自動獲得する Pre が、支持と反論の両方のデータを利用する Swi のみでは捉えられない、よりさまざまな支持や反論を獲得できていることを示唆している。

表 5.6 に、追加の調査として、Pol の有効性を確認するため、Crowdsourcing のリファレンスの情報に基づいてガイドを与えた場合の、開発データにおける Pol を用いる各発話生成手法の自動評価結果（perplexity）を示す。Baseline と比較して、Pol を利用する各発話生成手法の perplexity が全て低くなっている。この結果から、Pol を用いた場合でも、リファレンスと同じガイドを利用すれば、リファレンスの論拠がより生成されやすくなり、すなわち、Pol が支持や反論の生成に有効であることが確認できる。

表 5.7: 各発話生成手法が出力した主張および反論の論拠の人手評価結果（5段階のリッカート尺度）. +Pre+Pol は提案手法である. 人手で作成したリファレンス Human の評価値を除いて, 太字の数値は, 太字でない数値と有意な差があることを示す (Steel-Dwass の多重比較において, $p < .05$).

発話生成手法	Crowdsourcing		Artificial	
	支持	反論	支持	反論
Baseline	3.80	3.07	3.14	3.06
+Pre	3.89	3.22	3.73	3.49
+Pol	3.79	3.60	3.64	3.42
+Swi	3.75	3.21	3.55	3.16
+Pre+Swi	3.54	3.05	3.74	3.28
+Pol+Swi	3.81	3.62	3.79	3.79
+Pre+Pol(提案手法)	3.93²	3.76	3.91	3.91
+Pre+Pol+Swi	3.76	3.68	3.95	3.94
Human	4.35	3.89	NA	NA

5.4.4 人手評価

人手評価として, 各発話生成手法が出力した論拠が適切な支持または反論になっているかを調査した. 評価者として, ランサーズのクラウドソーシングプラットフォームを利用し, 10名の作業者が各出力を評価するようにした. 評価者は日本語を母国語とする作業者のみに制限し, その他の制限は設けなかった. 各評価者は, 各発話生成手法が出力した論拠が, 適切な支持または反論になっているかを評価した. 支持として生成された論拠を評価する際は適切な支持になっているかを評価し, 反論の場合は支持として不適切と判断するようにした. 作業者には教示として, 「出力発話が入力発話に対する支持 (反論) の理由を具体的に述べているか」を5段階のリッカート尺度 [43] で付与するように指示した. このとき, 出力発話が逆の立場 (支持なのに反論, または, 反論なのに支持) になっていたり, 入力発話と類似の内容 (単語) しか含んでいなかったり, 入力発話と無関係な内容になっている場合は, 支持および反論として不適切とするよう指示した. 実際に作業者に提示した教示は付録 5.6.2 に示す.

表 5.7 に, 各発話生成手法が出力した主張および反論の論拠の人手評価結果 (5段階のリッカート尺度) を示す. 表には, Crowdsourcing と Artificial からなる二種類の入力に対して生成された支持と反論の人手評価結果が示されている. この表から, 提案手法である +Pre+Pol が最も高い評価値を示しており, 手法の有効性が確認できる. 提案手法と Baseline を比較すると, Crowdsourcing における反論,

²例外的に, Crowdsourcing において, 提案手法である +Pre+Pol の評価値 (3.93) と +Pre の評価値 (3.89) の間には有意な差が見られなかった.

および、Artificialにおける支持と反論において大きく評価値が改善していることがわかる。この結果から、Crowdsourcingにおいて、Baselineは入力に関わらず支持として多い立場（賛成）の論拠を出力するよう学習されており、入力の立場を適切に考慮できていないことから、Crowdsourcingの反論において、および、入力の立場をバランスしたArtificialにおいて、評価値が低くなっていると考えられる。各発話生成手法、および、Humanの全てにおいて、支持よりも反論の評価値が低くなっていることから、支持と比較して反論を生成することがより難しい課題であるということが示唆される。この傾向は、表5.2に示した通り、支持は入力となる主張と類似する（同じ立場の）内容を論拠とすればよいのに対し、反論では、入力となる主張と異なる（逆の立場の）内容を論拠にする必要があり、よりさまざまな表現が可能であるためであると考えられる。

各素性を1種類のみ利用する発話生成手法である+Pre, +Pol, +SwiとBaselineを比較すると、概ね全ての評価値が改善していることから、各素性の有効性が確認できる。表5.5で示した自動評価結果同様、単に支持と反論のデータを併合して学習データの量を増加させる+Swiよりも+Preの方が評価値が大きく改善していることから、学習データを自動獲得する+Preの有効性が確認できる。+Pre同様、+Polについても、同様に+Swiと比較して大きく評価値が改善していることから、明示的に立場を考慮することが支持や反論の生成において重要であることが確認できる。

提案手法である+Pre+PolとさらにSwiも利用する+Pre+Pol+Swiを比較すると、Crowdsourcingにおいて、Swiを利用した方が支持と反論の両方で評価値が低いことがわかる。これは、支持と反論の両方を一つの発話生成モデルで同時に学習することが容易ではないことに起因すると考えられる。Swiを利用する手法では、学習データの量が支持と反論で合わせておよそ倍になるため、評価値が改善しているが、大規模に学習データを獲得するPreを利用する手法ではSwiを利用しなくても本研究で利用した発話生成モデルをfine-tuningする上で十分な規模のデータが収集できていると考えられる。支持と反論を同時に学習した場合、入力の表現と出力の表現の組み合わせは、支持と反論を個別に学習した場合と比較して多様になることから、学習が難しくなると考えられる。そのため、大規模な学習データを利用する提案手法（+Pre+Pol）ではSwiを利用しても評価値が改善しなかったと考えられる。この結果は、Polを利用しない+Preと+Pre+Swiを比較したとき、+Pre+Polと+Pre+Pol+Swi同様評価値の改善が見られないことから確認できる。以上の結果から、今回利用したBERTベースの発話生成手法においては、支持と反論は個別に学習する方がよい可能性がある。

表5.8に、ベースライン（Baseline）と提案手法（+Pre+Pol）の出力の例を示す。表には4種類の入力に対して生成された出力が大きく4つの行に分かれて示されている。上部の2種類の入力はベースラインと提案手法の両方で適切な支持や反論を生成できた例を示しており、下部の2種類の入力は提案手法のみで適切な支持と反論が生成できた例を示している。上部の2種類のような入力は、ベース

表 5.8: ベースライン (Baseline) と提案手法 (+Pre+Pol) の出力の例. 入力に続く括弧は、主張の立場を表す. 手法に続く括弧は、各手法が出力すべき論拠 (支持か反論か) を表す.

手法	生成された出力 (論拠)	評価値
入力 1 (賛成の主張)	オートバイが乗れたらいい。	NA
ベースライン (支持)	風が気持ちいい。	4.6
提案手法 (支持)	運転が楽しい。	4.5
ベースライン (反論)	事故が怖い。	4.7
提案手法 (反論)	バイクは危ない。	4.8
入力 2 (反対の主張)	たくあんは臭い。	NA
ベースライン (支持)	香りがいい。	1.9
提案手法 (支持)	匂いが苦手。	4.2
ベースライン (反論)	好き嫌いがあります。	3.4
提案手法 (反対の主張)	好みがあります。	3.8
入力 3 (賛成の主張)	競馬は楽しい。	NA
ベースライン (支持)	迫力があります。	4.7
提案手法 (支持)	予想が楽しい。	5.0
ベースライン (反論)	競馬の方が楽しい。	1.1
提案手法 (反論)	お金がかかります。	4.7
入力 4 (反対の主張)	伊勢海老はあまり好きではありません。	NA
ベースライン (支持)	種類がたくさんある。	1.5
提案手法 (支持)	匂いが苦手です。	4.7
ベースライン (反論)	好き嫌いがあります。	3.4
提案手法 (反対の主張)	味が美味しいです。	4.6

ラインの学習で利用した主張・論拠ペアデータで頻出の事例であり、そのような事例についてはベースラインでも適切な出力が得られている. 一方、下部の2種類のような頻度の少ない入力として、「競馬」のような頻出しない議論のトピックや、「好きではありません」のような頻出しない主張の表現では、ベースラインでは適切な支持や反論が生成されなかった.

ベースラインにおける典型的な誤りにおいて、反論として「競馬の方が楽しい」のように、議論のトピックを繰り返すような発話や、支持として「種類がたくさんある」のように、主張と合わない一般的な内容の発話が見られた. 前者の誤りは、発話生成モデルにおいて一般的に利用されるコピー機構 [23] に起因する誤りであると考えられる. 既存の一般的な発話生成モデルでは、出力される単語の確率分布に加えて、入力された単語のいずれか適切と考えられるものを出力として繰り返すようモデル化されており、入力中の単語や表現を繰り返しやすいという問題がある. 今回の例では、反論という立場を適切に考慮できていないことから、

表 5.9: 提案手法 (+Pre+Pol) で適切な支持や反論が生成できなかった出力の例. 入力に続く括弧は, 主張の立場を表す. 手法に続く括弧は, 各手法が出力すべき論拠 (支持か反論か) を表す.

手法	生成された出力 (論拠)	評価値
入力 5	マンションの見学は楽しい。	NA
ベースライン (支持)	歴史があります。	2.5
提案手法 (支持)	部屋がきれいになります	1.9
ベースライン (反論)	夏は暑い。	3.1
提案手法 (反論)	料金が低い	2.7
入力 6	靴はヒールが高い方がオシャレ。	NA
ベースライン (支持)	ストレス発散になります。	2.0
提案手法 (支持)	履き心地がいい	2.9
ベースライン (反論)	効果があります。	1.7
提案手法 (反論)	足が痛くなります。	3.9

この繰り返しの問題が起きていると考えられる. 後者の誤りは, dull response と呼ばれており, 学習データ中で一般的な応答が出力されやすい問題に起因すると考えられる [40]. 発話生成の問題は一般的にマルチリファレンス (正解となる出力が複数存在する) という問題であり, リファレンスの表現も多様であることから, 一般的な応答が出力されやすいよう学習が進む. この問題を抑制するため, 言語モデルを用いて出力のリランキングを行ったり [40], 学習された発話生成モデルから生成されやすい単語を明示的に抑制するような目的関数を発話生成モデルの学習時に設定する手法 [97, 41] が提案されている. これらの手法はベースラインの問題にも有効であると考えられるが, 支持や反論などの関係を明示的に考慮することができない. 提案手法では, 大規模な学習データの獲得による多様な表現の学習, および, 議論における立場の明示的な考慮による不適切な表現の生成の抑制ができていていると考えられる.

5.4.5 提案手法の誤り分析

表 5.9 に, 提案手法 (+Pre+Pol) で適切な支持や反論が生成できなかった出力の例を示す. 入力 5 については, 提案手法が生成した支持, 反論は共にマンションに関する話題であるものの, 入力された「見学が楽しい」という命題と無関係な内容が生成されており, 低い評価値が付与されている. また, 入力 6 については, 反論について適切な出力が生成されているものの, 支持については, 入力された「ヒールが高い」という命題に対して, 一般にヒールが高い靴は履き心地が悪いと考えられ, 矛盾する出力が生成されている. これらの出力については, 含意関係認識 [15] や矛盾検出 [84] が可能なモデルを用いて, 出力発話のフィルタリングを

することで部分的には解決できると考えられる。しかしながら、これらのモデルが必ずしも含意や矛盾を適切に判定できるとは限らないため、学習データの中で既知の情報（この例の場合はマンションや靴など）の情報に基づいて、荒い粒度から尤もらしい粒度まで発話を段階的に生成していく手法 [53] を検討することも重要だと考えている。

5.5 本章のまとめ

本章では、第4章までに得られた知見を間接的に利用して、議論対話システムの構築に取り組んだ。雑談対話システムにおいては、言外の情報の種類のうち、事実1および事実2の種類の伝達発話の自然さにおける評価値が最大となったため、言外の情報を考慮する雑談対話システムは、事実に基づく対話、かつ、今後人間との共同作業につながる実用性を伴う対話を扱うことが適切だと考え、議論対話システム構築の課題を設定した。深層学習ベースの発話生成モデルを議論の発話生成に利用し、支持または反論のペアを正確に発話生成モデルに学習させるために、ガイドを導入した。提案手法を人手評価した結果、5点満点の評価において、支持で3.93点、反論で3.76点となり、支持と反論の両方で評価値の有意な向上が確認できた。ベースラインと提案手法の間には、有意な差が見られた。この結果から、提案手法が論拠の生成に有効であることが確認できた。以上の結果から、言外の情報に基づく対話として、支持や反論の論拠を述べる議論対話において、本手法が有効であることが確認できた。本手法は部分的ではあるものの、言外の情報を扱うことができ、人間との共同作業につながる実用性があるものだと考えている。

5.6 本章の付録

5.6.1 主張・論拠ペアデータ収集のための教示と記述用フォーム

下記の回答ルールをよくお読みいただき、次のページより設問にお答えください。

【回答ルール】

【1】

はじめに、好きなもの、興味があるもの、趣味、嫌いなもの、など、思いついた事柄（「名詞」の単語）を3つお書きください。

どんなことでも結構ですが、「名詞」の単語に限定してください。

なお、できるだけ、人名や固有名詞は避けてください。

例：「イチロー」はNG、「メジャーリーグ」はOK

例：「村上春樹」はNG、「読書」はOK

【2】

上記で記述いただいた3つの名詞について、

他者と雑談をする想定で、議論（賛成派と反対派が分かれそうな話題）のきっかけになる言葉をお書きください。

できるだけ、実際にご自身が使うような話し言葉で記述してください。

また、「です」「ます」などの丁寧度で記述してください。

例：（「メジャーリーグ」）⇒「メジャーリーグが面白いですね」など

例：（「読書」）⇒「読書はつまらないです」など

【3】

「【2】」で話した言葉を聞いて、

賛同することを想定して、その理由を具体的に（※）にお書きください。

「そうだよな」の後に、続く理由をイメージしてください。

（※）「具体的」については、【1】で設定した言葉に関連する別の言葉を必ず入れてください。

（メジャーリーグ⇒「日本人選手」等）

例：（「メジャーリーグが面白いですね」）⇒「日本人選手が活躍していますからね」など

例：（「読書はつまらないです」）⇒「読書は目が疲れるから苦手です」など

【4】

「【2】」で話した言葉を聞いて、

反対する（賛同しない）ことを想定して、その理由を具体的に（※）にお書きください。

「そうは思わない」の後に、続く理由をイメージしてください。

（※）「具体的」については、【1】で設定した言葉に関連する別の言葉を必ず入れてください。

（読書⇒「ミステリー」等）

例：（「メジャーリーグが面白いですね」）⇒「日本のプロ野球のほうが面白いです」など

例：（「読書はつまらないです」）⇒「ミステリーが面白いです」など

【注意事項】

- 【1】～【4】すべての自由記述で、1つでも同じ文章を記入することはできません。
- 【2】～【4】の文章は、5文字以上とします。
- 【3】と【4】の理由の文章は、必ず「主語」＋「述語」で作成してください。
- すべて、日本語で記入してください。

図 5.6: 主張・論拠ペアデータ収集のために作業者に提示した教示

【2】1つ目の単語「**[Q1(1)]**」を使って、人との雑談のなかで議論（賛成派と反対派が分かれそうな話題）のきっかけになる言葉をお書きください。

例：「メジャーリーグ」の場合、「メジャーリーグが面白いですよ」など

【必須】

- できるだけ、実際にご自身が使うような話し言葉で記述してください。
- 「です」「ます」などの丁寧度で記述してください。

【3】前問で記述した「**[Q1s2(1)]**」を聞いて、賛同することを想定して、その理由を具体的にお書きください。

例：「メジャーリーグが面白いですよ」の場合、（そうだよね）「日本人選手が活躍していますからね」など

【必須】

『**[Q1s2(1)]**』

↓

「そうですね」

↓

（具体的な理由）

- 【1】で設定した言葉に関連する別の言葉を必ず入れてください。（メジャーリーグ⇒「日本人選手」等）
- 理由の文章は、必ず「主語」＋「述語」で作成してください。

【4】前問で記述した「**[Q1s2(1)]**」を聞いて、反対することを想定して、その理由を具体的にお書きください。

例：「メジャーリーグが面白いですよ」の場合、（そうは思わない）「日本のプロ野球のほうが面白いです」など

【必須】

『**[Q1s2(1)]**』

↓

「そうは思わないです」

↓

（具体的な理由）

- 【1】で設定した言葉に関連する別の言葉を必ず入れてください。（メジャーリーグ⇒「日本人選手」等）
- 理由の文章は、必ず「主語」＋「述語」で作成してください。

図 5.7: 主張・論拠ペアデータ収集における、主張、および、支持・反論となる論拠の記述フォーム

5.6.2 出力された論拠の評価に利用した教示

入力発話に対して、出力発話の支持（反論）らしさを評価してください。支持（反論）の条件は、出力発話が入力発話に対する支持（反論）の理由を具体的に述べているかどうかで判断します。支持（反論）になっているかどうかの確認は、入力発話と出力発話を「そうは思わない。なぜなら」（反論：「その通りだ。なぜなら」）でつなげることができるかで判断してください。出力発話が逆の立場（支持なのに反論、または、反論なのに支持）になっていたり、入力発話と類似の内容（単語）しか含んでいなかったり、入力発話と無関係な内容になっている場合は、支持（反論）らしさが低いと判断してください。入力発話や出力発話の内容に曖昧性がある場合は、出力発話を支持（反論）としても解釈することができれば、支持（反論）として判断する。

評価は5段階のリッカート尺度で行います。これは、下記の定義文に対する同意の度合を示します。

定義文 「出力発話は、入力発話の支持（反論）としてその理由を具体的に述べている。」

この定義文に対してどの程度同意できるかを、下記の5段階（1点～5点）のリッカート尺度で付与してください。

- (1) まったくそう思わない
- (2) そう思わない
- (3) どちらでもない
- (4) そう思う
- (5) 非常にそう思う

評価値を付与する際は、1点や5点のみ、2点から4点のみなど、一部の評価値のみを使うのではなく、1点から5点の全体を使うようにしてください。

第6章 結論

本章では、結論として本論文のまとめ、および、本論文の成果における今後の展望と今後の課題について述べる。

6.1 本論文のまとめ

本論文は、ユーザとの高度な対話が可能な雑談対話システムを構築するために、雑談対話において発話に明示されていなくとも人間が理解できる重要な情報を言外の情報と定義し、言外の情報のデータ収集と類型化、有用性の検証、言外の情報に基づく対話システムの構築に取り組み、その成果をまとめたものである。高度な対話では、参加者が対話の内容の理解を積み上げていく必要があるため、その理解を言外の情報として明らかにし、雑談対話システムに応用することを狙った。これまで、対話研究、および、自然言語処理研究において、言外の情報に類する情報を扱った数多くの研究が報告されている。しかしながら、言外の情報は抽象的な情報であるため、特定の状況で言外の情報が伝達される枠組みを理論的に説明する試みはあるものの、雑談対話において具体的にどのような言外の情報が伝達されているのか、また、伝達された言外の情報が雑談対話システムにどう応用可能なかは明らかにされてこなかった。そこで本研究では、言外の情報をデータとして収集することがその類型化や応用を検討するために重要だと考え、多数の作業者に言外の情報を記述させ、記述された言外の情報をまとめ上げるというアプローチを取ることで、言外の情報を定量的に扱うことを試みた。

第1章では、研究背景、先行研究課題と本研究の位置付け、本研究のアプローチについて述べた。

第2章では、対話の理論研究、対話システムの類型、代表的なタスク指向型対話システムおよび雑談対話システム、深層学習に基づく雑談対話システムの発展、対話システムに関するデータセットと評価について述べた。

第3章では、対話において人間が発話から理解する情報のうち、発話の内容とは異なる、明示されていなくとも人間が理解する情報、これを「言外の情報」(perceived information)と定義し、そのデータ収集と類型化を行った。まず、言外の情報のデータ収集を行った。収集手順として、まず、言外の情報を記述する対象となる対話テキストを用意した。次に、複数の作業者が対話テキスト中の各発話に対して言外の情報を記述することで、言外の情報を収集した。一般的かつ多様な言外

の情報を網羅的に収集するために、作業対象として大規模な雑談対話コーパスを利用し、調査を行うのに十分な規模と考えられる、43 名の話者による 1,257 発話を含む 30 対話分の雑談対話テキストを利用した。これらのテキストに対し、6 名の作業者が独立に全テキスト中の発話に対して言外の情報を記述することで、計 11,533 事例の言外の情報を収集した。次に、収集された言外の情報を用いて、複数の作業者が人手で類似する事例をまとめ上げていくことで、言外の情報を類型化した。類型を作成する際には、特定の作業者による恣意的な類型ではなく、一般性を持った類型が作成されるよう、複数の作業者が類型化を行い、それらを統合するようにした。最終的に、レベル 1 からレベル 3 まで、階層的に言外の情報の類型が列挙された。最も細かい類型は 9 種類（願望 1、願望 2、信念 1、信念 2、属性、環境、行動、事実 1、事実 2）となった。作成された類型の妥当性を評価するために、言外の情報に対して人手でアノテーションを行い、作業者間一致率を調査した。その結果、類型の単純一致率で 0.75、カッパ値で 0.69 と高い一致率でアノテーションが可能な類型であることを確認した。

第 4 章では、雑談対話システムにおける言外の情報の有用性を検証するために、最も基本的な方法として、理解した言外の情報をそのままシステムの発話として応答したとき、ユーザにどのような影響を及ぼすのか、また、雑談対話において有用な言外の情報が存在するのかを検証した。具体的には、雑談対話システムが発話から理解しうるさまざまな種類の情報をユーザに伝達した際の影響について分析を行い、自然さを損ねることなくシステムが発話可能な言外の情報を明らかにした。検証手順として、言外の情報を伝達するシステム発話を人手で作成、評価することで、異なる種類の言外の情報を伝達した際のシステム発話の自然さを被験者実験により調査した。言外の情報の類型として、最も細かい言外の情報の類型（レベル 3）を利用し、比較対象として、雑談対話システムによる応答（自動応答発話）、ユーザ発話を繰り返す応答（繰り返し発話）、人手で作成した適切な応答（人手応答発話）を用意した。人手評価の結果、言外の情報の伝達発話は全体としては繰り返し発話や人手応答発話と比べて自然さが低く評価されたものの、個別の類型としては大きく三群（低評価群、中評価群、高評価群）に分けられることがわかり、これら 3 つの群の間には、有意な差が見られた。低評価群に該当する対話中のユーザの一時的な状態（例えば、対話中のユーザの考え）はシステム発話として自然さの評価値が低く、中評価群に該当する一方でユーザ属性や、高評価群に該当するユーザ自身と直接関係のない事実に関する情報はシステム発話として自然さを損ねることなく伝達することができることが明らかになった。

第 5 章では、第 4 章までに得られた知見を間接的に利用して、議論対話システムの構築に取り組んだ。雑談対話システムにおいては、言外の情報の種類のうち、事実 1 および事実 2 の種類の伝達発話の自然さにおける評価値が最大となったため、言外の情報を考慮する雑談対話システムは、事実に基づく対話、かつ、今後人間との共同作業につながる実用性を伴う対話を扱うことが適切だと考え、議論対話システム構築の課題を設定した。深層学習ベースの発話生成モデルを議論の発

話生成に利用し、支持または反論のペアを正確に発話生成モデルに学習させるために、ガイドを導入した。ガイド付きの学習データ（主張・論拠ペアデータ）は、二種類の方法（自動収集と人手収集）で作成した。自動収集では、大規模テキストから自動的に主張・論拠ペアデータを獲得する。品質は完全ではないものの、さまざまなトピックに関するペアを獲得することができた。もう一方の人手収集では、人手で主張・論拠ペアデータを書き下す。規模は限られるものの、高品質なペアを収集することができた。提案手法を人手評価した結果、5点満点の評価において、支持で3.93点、反論で3.76点となり、支持と反論の両方で評価値の有意な向上が確認できた。ベースラインと提案手法の間には、有意な差が見られた。この結果から、提案手法が論拠の生成に有効であることが確認できた。以上の結果から、言外の情報に基づく対話として、支持や反論の論拠を述べる議論対話において、本手法が有効であることが確認できた。本手法は部分的ではあるものの、言外の情報を扱うことができ、人間との共同作業につながる実用性があるものだと考えている。

本研究は、雑談対話における言外の情報を明示的に扱うことを試みた最初の研究である。現在の雑談対話システムは、人間が容易に理解可能な言外の情報を扱うことができないため、人間のような協調的な振舞いができずシステムの応用範囲が限定されている。人間同士の雑談は、関係構築、傾聴による承認、相談など、人間社会において非常に多様かつ重要な役割を持っている。これらの役割を達成する対話は、対話の参加者が言外の情報を適切に汲み取り、それに合った適切な振舞いをすることで達成される高度な共同作業である。今後雑談対話システムが言外の情報を適切に扱うことが可能になれば、既存の雑談対話システムの限界を越えた高度な共同作業が可能となり、音声操作や情報案内に留まらない多様な対話の実現可能となる。将来的には、システム設計者やユーザに指示された通りに振る舞うシステムだけでなく、ユーザとの間で信頼関係が築かれたパートナーとして、多様な状況でユーザを支えるシステムも実現されるだろう。本研究で明らかになった言外の情報に関する知見は、今後そういった既存の限界を越えた対話システムを構築する際の一助になると考えている。

6.2 今後の課題

今後の展望として、本技術を既存の雑談対話システムに導入することが考えられる。対象のシステムとして、例えば、システムがユーザから主体的に話を聞くことでインタビューや傾聴対話を行う雑談対話システムが挙げられる。傾聴対話の研究では、システムがユーザに対して質問するだけでなく、ユーザの回答に対して適切な応答を返すことで、対話が円滑になることが報告されている。ユーザの発話、例えば、「横浜に住んでいます」という発話）に対して、その内容を支持する発話、「海が綺麗だね」を返すことで、ユーザはシステムに理解されている印象を持ち、対話が円滑になる可能性がある。また、別のシステムとして、特定の

個性を持った雑談対話システムへの導入も考えられる。既存のアニメキャラクターや実在の人物はあらかじめその個性として、好きなものや嫌いなものが明らかになっていることが多い。そのため、その個性と矛盾しない形で発話を生成するために支持や反論を使い分けることで、一貫性を持った対話が可能になると考えられる。

既存の雑談対話システムのみならず、より高度な対話として、人間との共同作業を行うような対話システムへの応用も考えられる。例えば、個人の意思決定支援を行うために、各個人が関心を持つトピック（例えば、「今日の晩ご飯で寿司とラーメンのどちらを食べるか」）について議論を行うことができる対話システムへの応用が考えられる。ユーザの意見に対して、支持の意見や反論の意見をシステムが多様に生成することで、ユーザはさまざまな観点から気軽に意思決定をすることができるようになる。また、アイディア創発や教育につながるような、創造的な雑談を行うシステムへの応用も考えられる。具体的には、特定のトピックについてアイディアを出したり（例えば、企業における商品企画など）、トピックへの理解を深めたり（例えば、学校におけるディベートなど）するシーンが挙げられる。そのトピックについて十分な知識を持つ対話システムが意見を生成したり、その意見の理由を述べることであれば、ユーザは特定のトピックについて十分な知識がなくとも多様なアイディアを考えられたり、トピックに対する理解や議論の方法を学べたりする可能性があると考えられる。

今後の課題として、これらの応用を実現するためには、まず、過去文脈を考慮した議論対話システムの実現が必要である。2.4で述べたように、一部例外的に大規模な対話データに基づいて文脈を考慮する対話システムが提案され始めたものの、多くの雑談対話システムは一問一答で構築されており、文脈を考慮して対話を進める手法は確立されていない。文脈も考慮可能な雑談対話システムを構築するためには所望のドメインにおける大規模な対話データが必要となるため、今後のアプローチとして、大規模な議論対話を収集し、大規模対話モデルを構築する手法が考えられる。しかし、議論対話システムを学習する上で十分な規模の対話データを収集できるかどうかは明らかではないため、少ない対話データのみから効率的に文脈を考慮する HRED や VHRED[76, 77] のような学習手法を検討することも重要だと考えられる。

本研究では議論対話を対象に、支持や反論を生成するという目的を設定したが、雑談が持つ多様な目的に合わせて、システムはどのような発話を生成すべきかというプランニングの手法を確立することも必要だと考えられる。例えば、相手を説得するような場合であれば、システムは説得力のある対話を実現することが望ましい。また、アイディア創発を支援するような場合であれば、システムはユーザが思い付きにくい多様なアイディアが出現するように対話を実現することが望ましい。これらの個々の目的を達成するために、対話システムが状況に応じてどのような発話を生成すべきかは明らかではないため、これをプランニングの問題として解決する必要がある。近年タスク対話における有用性が報告されている強

化学習ベースのタスク指向型対話システム [42] は、この課題において、目的ごとに対話データを収集でき、かつ、対話の目的の達成度合いを適切に評価することができれば、有用なアプローチだと考えられる。

本研究が言外の情報をそのまま発話するアプローチを取っているのに対し、言外の情報を考慮した発話を選択したり生成することも、雑談対話の性能を向上させる上で有効な手段だと考えられる。例えば、聞き手への願望を表す願望2の言外の情報を考慮して、ユーザが褒められたいということをシステムが理解した際は、ユーザを褒める発話を生成したり、ユーザが話題を変えたいと思っている場合は、新しい話題を導入する発話を生成することが考えられる。2.4節で述べたように、深層学習ベースの雑談対話システムにおいて、発話を外部から制御することは難しいことが報告されているが、例えば発話を条件付けする手法である Conditional Variational AutoEncoder[107] や Unlikelihood Training[41] は、本課題を解決するための有望なアプローチだと考えられる。

謝辞

本論文は、筆者が日本電信電話株式会社NTTメディアインテリジェンス研究所、および、筑波大学大学院システム情報工学研究科において行った研究成果をまとめたものです。本研究を遂行するにあたり、また、本論文を執筆するにあたり、数多くの方々から多大なるご指導を賜りました。特にお世話になった方々についてここに記し、深謝の意を表します。

指導教員である筑波大学 宇津呂武仁教授には、本研究の内容、および、本論文の構成について丁寧なご指導を賜りました。また、副指導教員である矢野博明教授、星野准一准教授、副査をお引き受けいただいた丸山勉教授、山本幹雄教授には、お忙しいなか本論文の審査にご協力いただき、本論文の内容に関する有用なご助言を頂きました。

NTTメディアインテリジェンス研究所にて本研究を遂行するにあたり、名古屋大学 東中竜一郎教授（および、NTTメディアインテリジェンス研究所・NTTコミュニケーション科学基礎研究所客員上席特別研究員）には多大なるご指導を賜りました。本研究は、対話の中で人間が理解する情報を工学的に明らかにしようという新しい取り組みであり、対話システム研究の第一人者である東中教授のご指導なくしては達成し得ない内容であったと確信しております。また、研究の内容のみならず、問題を解決するための心構えや粘り強さの重要性をご教授いただき、私自身が大きく成長できたと実感しています。

NTT研究所にてお世話になった方々のお名前はここには書き切ることができませんが、特に、元グループリーダーの松尾義博氏、および、富田準二博士、現プロジェクトマネージャの青野裕司博士、現グループリーダーの杵渕哲也氏には本研究の遂行をご理解、ご支援いただき、有益なご助言を頂きました。また、本研究の内容について、貞光九月博士、西田京介特別研究員、小林のぞみ博士、石井亮博士、杉山弘晃博士、片山太一氏、大塚淳史博士には有益なご助言やご議論の機会を頂きました。

最後に、いつも暖かく私を支え、応援してくれた家族に感謝します。

参考文献

- [1] Daniel Adiwardana, Minh-Thang Luong, David R So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, et al. Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977*, 2020.
- [2] James Allen. *Recognizing Intentions from Natural Language Utterances*. MIT Press, 1983.
- [3] John Langshaw Austin. *How to Do things with words*. Oxford University Press, 1962.
- [4] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.
- [5] Rafael E. Banchs and Haizhou Li. IRIS: a chat-oriented dialogue system based on the vector space model. In *Proc. ACL*, pp. 37–42, 2012.
- [6] Yonatan Bilu, Daniel Hershcovich, and Noam Slonim. Automatic claim negation: why, how and when. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pp. 84–93, 2015.
- [7] Michael Bratman. *Intention, plans, and practical reason*, Vol. 10. Harvard University Press Cambridge, MA, 1987.
- [8] Penelope Brown and Stephen C. Levinson. *Politeness: Some universals in language usage*, Vol. 4. Cambridge university press, 1987.
- [9] Harry Bunt, Jan Alexandersson, Jean Carletta, JaeWoong 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 *Proc. LREC*, 2010.
- [10] Nathanael Chambers and Dan Jurafsky. Unsupervised learning of narrative event chains. In *Proc. ACL*, 2008.

- [11] Sihao Chen, Daniel Khashabi, Wenpeng Yin, Chris, Callison-Burch, and Dan Roth. Seeing things from a different angle: Discovering diverse perspectives about claims. In *Proc. NAACL-HLT*, pp. 542–557, 2019.
- [12] Herbert H. Clark. *Using Language*. Cambridge University Press, 1996.
- [13] Gonalo M. Correia and Andr  F. T. Martins. A simple and effective approach to automatic post-editing with transfer learning. In *Proc. ACL*, pp. 3050–3056, 2019.
- [14] Ido Dagan, Oren Glickman, and Bernardo Magnini. The pascal recognising textual entailment challenge. In *Proc. of the Second PASCAL Challenges Workshop on Recognising Textual Entailment*, pp. 1–15, 2006.
- [15] Ido Dagan, Oren Glickman, and Bernardo Magnini. The pascal recognising textual entailment challenge. In *Proc. of the Second PASCAL Challenges Workshop on Recognising Textual Entailment*, pp. 1–15, 2006.
- [16] Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. A computational approach to politeness with application to social factors. In *Proc. ACL*, pp. 250–259, 2013.
- [17] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proc. NAACL-HLT*, pp. 4171–4186, 2019.
- [18] Meyer Dwass. Some k-sample rank-order tests. *Contributions to probability and statistics*, pp. 198–202, 1960.
- [19] L. J. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, Vol. 76, No. 5, pp. 378–382, 1971.
- [20] Shun-ya Fukunaga, Hitoshi Nishikawa, Takenobu Tokunaga, Hikaru Yokono, and Tetsuro Takahashi. Analysis of implicit conditions in database search dialogues. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.
- [21] John J Godfrey, Edward C Holliman, and Jane McDaniel. SWITCHBOARD: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, Vol. 1, pp. 517–520. IEEE Computer Society, 1992.

- [22] Herbert Paul Grice. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and Semantics*, Vol. 3: Speech Acts, pp. 41–58. Academic Press, 1975.
- [23] Jiatao Gu, Zhengdong Lu, Hang Li, and Victor OK Li. Incorporating copying mechanism in sequence-to-sequence learning. *arXiv preprint arXiv:1603.06393*, 2016.
- [24] Swati Gupta, Marilyn A. Walker, and Daniela M. Romano. How rude are you?: Evaluating politeness and affect in interaction. In *Proc. International Conference on Affective Computing and Intelligent Interaction*, pp. 203–217, 2007.
- [25] Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. Semeval-2018 task 12: The argument reasoning comprehension task. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pp. 763–772, 2018.
- [26] Christopher Hidey and Kathleen McKeown. Fixed that for you: Generating contrastive claims with semantic edits. In *Proc. NAACL-HLT*, pp. 1756–1767, 2019.
- [27] Ryuichiro Higashinaka, Kenji Imamura, Toyomi Meguro, Chiaki Miyazaki, Nozomi Kobayashi, Hiroaki Sugiyama, Toru Hirano, Toshiro Makino, and Yoshihiro Matsuo. Towards an open domain conversational system fully based on natural language processing. In *Proc. COLING*, pp. 928–939, 2014.
- [28] Ryuichiro Higashinaka, Kenji Imamura, Toyomi Meguro, Chiaki Miyazaki, Nozomi Kobayashi, Hiroaki Sugiyama, Toru Hirano, Toshiro Makino, and Yoshihiro Matsuo. Towards an open domain conversational system fully based on natural language processing. In *Proc. COLING*, pp. 928–939, 2014.
- [29] Ryuichiro Higashinaka, Masahiro Mizukami, Hidetoshi Kawabata, Emi Yamaguchi, Noritake Adachi, and Junji Tomita. Role play-based question-answering by real users for building chatbots with consistent personalities. In *Proc. ACL*, 2018.
- [30] Ryuichiro Higashinaka, Kazuki Sakai, Hiroaki Sugiyama, Hiromi Narimatsu, Tsunehiro Arimoto, Takaaki Fukutomi, Kiyooki Matsui, Yusuke Ijima, Hiroaki Ito, Shoko Araki, Yuichiro Yoshikawa, Hiroshi Ishiguro, and Yoshihiro Matsuo. Argumentative dialogue system based on argumentation structures. In *Proc. SEMDIAL*, pp. 154–155, 2017.

- [31] Toru Hirano, Nozomi Kobayashi, Ryuichiro Higashinaka, Toshiro Makino, and Yoshihiro Matsuo. User information extraction for personalized dialogue systems. In *Proc. SEMDIAL*, pp. 67–76, 2015.
- [32] Toru Hirano, Nozomi Kobayashi, Ryuichiro Higashinaka, Toshiro Makino, and Yoshihiro Matsuo. User information extraction for personalized dialogue systems. In *Proc. SEMDIAL*, pp. 67–76, 2015.
- [33] Xinyu Hua and Lu Wang. Neural argument generation augmented with externally retrieved evidence. In *Proc. ACL*, pp. 219–230, 2018.
- [34] Koji Inoue, Divesh Lala, Kenta Yamamoto, Shizuka Nakamura, Katsuya Takanashi, and Tatsuya Kawahara. An attentive listening system with android ERICA: Comparison of autonomous and WOZ interactions. In *Proc. SIGDIAL*, pp. 118–127, 2020.
- [35] Dan Jurafsky. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual. 2015.
- [36] Seokhwan Kim, Michel Galley, Chulaka Gunasekara, Sungjin Lee, Adam Atkinson, Baolin Peng, Hannes Schulz, Jianfeng Gao, Jinchao Li, Mahmoud Adada, et al. The eighth dialog system technology challenge. *arXiv preprint arXiv:1911.06394*, 2019.
- [37] Yonghee Kim, Jeessoo Bang, Junhwi Choi, Seonghan Ryu, Sangjun Koo, and Gary Geunbae Lee. User information extraction for personalized dialogue systems. In *Proc. Workshop on Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction*, 2014.
- [38] Nozomi Kobayashi, Kentaro Inui, and Yuji Matsumoto. Extracting aspect-evaluation and aspect-of relations in opinion mining. In *Proc. EMNLP-CoNLL*, pp. 1065–1074, 2007.
- [39] Taku Kudo and John Richardson. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proc. EMNLP*, pp. 66–71, 2018.
- [40] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055*, 2015.

- [41] Margaret Li, Stephen Roller, Ilia Kulikov, Sean Welleck, Y-Lan Boureau, Kyunghyun Cho, and Jason Weston. Don't say that! making inconsistent dialogue unlikely with unlikelihood training. *arXiv preprint arXiv:1911.03860*, 2019.
- [42] Xiujun Li, Yun-Nung Chen, Lihong Li, Jianfeng Gao, and Asli Celikyilmaz. End-to-end task-completion neural dialogue systems. *arXiv preprint arXiv:1703.01008*, 2017.
- [43] Rensis Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.
- [44] Marco Lippi and Paolo Torroni. Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology*, Vol. 16, No. 10, 2016.
- [45] Diane J. Litman and James F. Allen. A plan recognition model for subdialogues in conversations. *Cognitive Science*, Vol. 11, pp. 163–200, 1987.
- [46] Chia-Wei Liu, Ryan Lowe, Iulian Serban, Mike Noseworthy, Laurent Charlin, and Joelle Pineau. How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. In *Proc. EMNLP*, pp. 2122–2132, 2016.
- [47] Ryan Lowe, Michael Noseworthy, Iulian Vlad Serban, Nicolas Angelard-Gontier, Yoshua Bengio, and Joelle Pineau. Towards an automatic turing test: Learning to evaluate dialogue responses. In *Proc. ACL*, pp. 1116–1126, 2017.
- [48] Chunling Ma, Helmut Prendinger, and Mitsuru Ishizuka. Emotion estimation and reasoning based on affective textual interaction. In *Proc. Affective Computing and Intelligent Interaction*, 2005.
- [49] Toyomi Meguro, Ryuichiro Higashinaka, Yasuhiro Minami, and Kohji Dohsaka. Evaluation of listening-oriented dialogue control rules based on the analysis of hmms. In *Proc. Interspeech*, 2011.
- [50] Toyomi Meguro, Yasuhiro Minami, Ryuichiro Higashinaka, and Kohji Dohsaka. Learning to control listening-oriented dialogue using partially observable markov decision processe. *ACM Transactions on Speech and Language Processing*, Vol. 10, No. 15, pp. 1–15, 2014.

- [51] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [52] Ashutosh Modi, Tatjana Anikina, Simon Ostermann, and Manfred Pinkal. InScript: Narrative texts annotated with script information. In *Proc. LREC*, 2016.
- [53] Amit Moryossef, Yoav Goldberg, and Ido Dagan. Step-by-step: Separating planning from realization in neural data-to-text generation. *arXiv preprint arXiv:1904.03396*, 2019.
- [54] Nasrin Mostafazadeh, Nathanael Chambers, Xiaodong He, Devi Parikh, Dhruv Batra, Lucy Vanderwende, Pushmeet Kohli, and James Allen. A corpus and evaluation framework for deeper understanding of commonsense stories. In *Proc. NAACL-HLT*, 2016.
- [55] Tong Niu and Mohit Bansal. Polite dialogue generation without parallel data. *Transactions of the Association for Computational Linguistics*, Vol. 6, pp. 373–389, 2018.
- [56] Kanako Onishi and Takeshi Yoshimura. Casual conversation technology achieving natural dialog with computers. *NTT DOCOMO Technical Journal*, Vol. 15, No. 4, pp. 16–21, 2014.
- [57] Kishore Papineni, Salim Roukos, Todd Ward, John Henderson, and Florence Reeder. Corpus-based comprehensive and diagnostic MT evaluation: initial arabic, chinese, french, and spanish results. In *Proceedings of the second international conference on Human Language Technology Research*, pp. 132–137, 2002.
- [58] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proc. ACL*, pp. 311–318, 2002.
- [59] ChaeHun Park, Wonsuk Yang, and Jong C. Park. ArgDiver: Generating sentential arguments from diverse perspectives on controversial topic. In *Proceedings of the 2nd Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda*, pp. 56–65, 2019.
- [60] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference*

on empirical methods in natural language processing (EMNLP), pp. 1532–1543, 2014.

- [61] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*, 2018.
- [62] Yada Pruksachatkun, Jason Phang, Haokun Liu, Phu Mon Htut, Xiaoyi Zhang, Richard Yuanzhe Pang, Clara Vania, Katharina Kann, and Samuel R. Bowman. Intermediate-task transfer learning with pretrained language models: When and why does it work? In *Proc. ACL*, pp. 5231–5247, 2020.
- [63] Niklas Rach, Wolfgang Minker, and Stefan Ultes. Utilizing argument mining techniques for argumentative dialogue systems. In *Proc. IWSDS*, pp. 1–12, 2018.
- [64] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.
- [65] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. In *arXiv preprint arXiv:1910.10683*, pp. 1–67, 2020.
- [66] Geetanjali Rakshit, Kevin K. Bowden, Lena Reed, Amita Misra, and Marilyn Walker. Debbie, the debate bot of the future. In *Proc. IWSDS*, pp. 1–6, 2017.
- [67] Hannah Rashkin, Maarten Sap, and Emily Allaway Noah A. Smith Yejin Choi. Event2Mind: Commonsense inference on events, intents, and reactions. In *Proc. ACL*, pp. 463–473, 2018.
- [68] Alan Ritter, Colin Cherry, and William B. Dolan. Data-driven response generation in social media. In *Proc. EMNLP*, pp. 583–593, 2011.
- [69] Hiroaki Sugiyama Toshiro Makino Yoshihiro Matsuo Ryuichiro Higashinaka, Toyomi Meguro. On the difficulty of improving hand-crafted rules in chat-oriented dialogue systems. In *Proc. APSIPA*, pp. 1014–1018, 2015.
- [70] Kazuki Sakai, Ryuichiro Higashinaka, Yuichiro Yoshikawa, Hiroshi Ishiguro, and Junji Tomita. Introduction method for argumentative dialogue using paired question-answering interchange about personality. In *Proc. SIGDIAL*, pp. 70–79, 2018.

- [71] Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatgula, Nicholas Lourie, Hannah Rashkin, Brendan Roof Noah A. Smith, and Yejin Choia. ATOMIC: An atlas of machine commonsense for if-then reasoning. In *Proc. AAAI*, pp. 3027–3035, 2019.
- [72] Misa Sato, Kohsuke Yanai, Toshihiko Yanase, Toshinori Miyoshi, Makoto Iwayama, Qinghua Sun, and Yoshiki Niwa. End-to-end argument generation system in debating. In *Proceedings of ACL-IJCNLP 2015 System Demonstrations*, pp. 109–114, 2015.
- [73] Roger C. Schank and Robert P. Abelson. *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Lawrence Erlbaum, 1977.
- [74] John R. Searle. *Speech acts: An essay in the philosophy of language*. Cambridge university press, 1969.
- [75] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*, 2015.
- [76] Iulian V Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. Building end-to-end dialogue systems using generative hierarchical neural network models. *arXiv preprint arXiv:1507.04808*, 2015.
- [77] Iulian Vlad Serban, Alessandro Sordoni, Ryan Lowe, Laurent Charlin, Joelle Pineau, Aaron Courville, and Yoshua Bengio. A hierarchical latent variable encoder-decoder model for generating dialogues. *arXiv preprint arXiv:1605.06069*, 2016.
- [78] Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proc. EMNLP*, pp. 1631–1642, 2013.
- [79] Christian Stab and Iryna Gurevych. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, Vol. 43, No. 3, 2017.
- [80] Naman Goyal Da Ju Mary Williamson Yinhan Liu Jing Xu Myle Ott Kurt Shuster Eric M. Smith Y-Lan Boureau Jason Weston Stephen Roller, Emily Dinan. Recipes for building an open-domain chatbot. 2020.

- [81] Hiroaki Sugiyama, Toyomi Meguro, Ryuichiro Higashinaka, and Yasuhiro Minami. Large-scale collection and analysis of personal question-answer pairs for conversational agents. In *International Conference on Intelligent Virtual Agents*, pp. 420–433. Springer, 2014.
- [82] Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. Lstm neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association*, 2012.
- [83] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, Vol. 27, pp. 3104–3112, 2014.
- [84] Yu Takabatake, Hajime Morita, Daisuke Kawahara, Sadao Kurohashi, Ryuichiro Higashinaka, and Yoshihiro Matsuo. Classification and acquisition of contradictory event pairs using crowdsourcing. In *Proc. the 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pp. 99–107, 2015.
- [85] Alexandre Tamborrino, Nicola Pellicano, Baptiste Pannier, Pascal Voitot, and Louise Naudin. Pre-training is (almost) all you need: An application to commonsense reasoning. In *Proc. ACL*, pp. 3878–3887, 2020.
- [86] Stephen Edelston Toulmin. *The Uses of Argument*. Cambridge University Press, 1958.
- [87] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, Vol. 30, pp. 5998–6008, 2017.
- [88] Oriol Vinyals and Quoc Le. A neural conversational model. In *Proc. ICML Deep Learning Workshop*, 2015.
- [89] Oriol Vinyals and Quoc Le. A neural conversational model. In *Proc. ICML Deep Learning Workshop*, pp. 1–7, 2015.
- [90] Marilyn A. Walker, Janet E. Cahn, and Stephen J. Whittaker. Improvising linguistic style: social and affective bases for agent personality. In *Proc. the first international conference on Autonomous agents*, pp. 96–105, 1997.
- [91] Richard S. Wallace. *The Anatomy of A.L.I.C.E.* A.L.I.C.E. Artificial Intelligence Foundation, Inc., 2004.

- [92] Douglas Walton. *Methods of argumentation*. Cambridge University Press, 2013.
- [93] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pp. 353–355, 2018.
- [94] William Yang Wang, Samantha Finkelstein, Amy Ogan, Alan W Black, and Justine Cassell. "love ya, jerkface": Using sparse log-linear models to build positive and impolite relationships with teens. In *Proc. SIGDIAL*, pp. 20–29, 2012.
- [95] Lilian D. A. Wanzare, Alessandra Zarcone, Stefan Thater, and Manfred Pinkal. DeScript: A crowdsourced database for the acquisition of high-quality script knowledge. In *Proc. LREC*, 2016.
- [96] Joseph Weizenbaum. ELIZA-a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, Vol. 9, No. 1, pp. 36–45, 1966.
- [97] Sean Welleck, Ilia Kulikov, Stephen Roller, Emily Dinan, Kyunghyun Cho, and Jason Weston. Neural text generation with unlikelihood training. *arXiv*, pp. arXiv–1908, 2019.
- [98] Jason D. Williams, Antoine Raux, Deepak Ramachandran, and Alan Black. The dialog state tracking challenge. In *Proc. SIGDIAL*, 404–413, 2013.
- [99] Xianchao Wu, Kazushige Ito, Katsuya Iida, Kazuna Tsuboi, and Momo Klyen. りんな: 女子高生人工知能. 言語処理学会第 22 回年次大会発表論文集, 2016.
- [100] Koichiro Yoshino, Shinsuke Mori, and Tatsuya Kawahara. Spoken dialogue system based on information extraction using similarity of predicate argument structures. In *Proceedings of the SIGDIAL 2011 Conference*, pp. 59–66, 2011.
- [101] Chen Zhang and Joyce Y. Chai. What do we know about conversation participants: experiments on conversation entailment. In *Proc. SIGDIAL*, pp. 206–215, 2009.
- [102] Chen Zhang and Joyce Y. Chai. Towards conversation entailment: an empirical investigation. In *Proc. EMNLP*, pp. 756–766, 2010.

- [103] Muyu Zhang, Bing Qin, Ting Liu, and Mao Zheng. Triple based background knowledge ranking for document enrichment. In *Proc. COLING*, pp. 917–927, 2014.
- [104] Muyu Zhang, Bing Qin, Mao Zheng, Graeme Hirst, and Ting Liu. Encoding distributional semantics into triple-based knowledge ranking for document enrichment. In *Proc. ACL/IJCNLP*, pp. 524–533, 2015.
- [105] Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. Personalizing dialogue agents: I have a dog, do you have pets too? *arXiv preprint arXiv:1801.07243*, 2018.
- [106] Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, Vol. 46, No. 1, pp. 53–93, 2020.
- [107] Xianda Zhou and William Yang Wang. MojiTalk: Generating emotional responses at scale. In *Proc. ACL*, pp. 1128–1137, 2018.
- [108] 井上昂治, 河原達也. 自律型アンドロイド Erica のための音声対話システム. pp. 21–24, 2015.
- [109] 宇佐美まゆみ. ポライトネス理論の展開 (4) b&l のポライトネス理論 (2) 「フェイス」という概念と 「フェイス侵害度見積もりの公式」. *言語*, Vol. 31, No. 4, pp. 96–101, 2002.
- [110] 岡本真一郎. 情報への関与と文末形式. *心理学評論*, Vol. 39, No. 2, pp. 168–204, 1996.
- [111] 河原達也. アンドロイド ERICA による人間レベルの音声対話. 第 84 回言語・音声理解と対話処理研究会, pp. 27–32, 2018.
- [112] 石崎雅人, 伝康晴. 談話と対話 (言語と計算). 東京大学出版会, 2001.
- [113] 鹿野清宏, Tobias Cincarek, 竹内翔大, 川波弘道, 李晃伸, 西村竜一, 猿渡洋. 音声情報案内システム「たけまるくん」の運用. 情報処理学会第 70 回全国大会, pp. 5:335–5:336, 2008.
- [114] 小磯花絵, 石本祐一, 菊池英明, 坊農真弓, 坂井田瑠衣, 渡部涼子, 田中弥生, 伝康晴. 大規模日常会話コーパスの構築に向けた取り組み-会話収録法を中心に-. 第 74 回言語・音声理解と対話処理研究会, 第 5 巻, pp. 37–42, 2015.
- [115] 神尾昭雄. 情報のなわ張り理論 – 言語の機能的分析. 大修館書店, 1990.
- [116] 神尾昭雄. 続・情報のなわ張り理論. 大修館書店, 2002.

- [117] 杉山弘晃, 成松宏美, 水上雅博, 有本庸浩, 千葉祐弥, 目黒豊美, 中嶋秀治. Transformer encoder-decoder モデルによる趣味雑談システムの構築. 言語・音声理解と対話処理研究会, pp. 104–109, 2020.
- [118] 石黒浩. アンドロイドによるトータルチューリングテストの可能性. 人工知能, Vol. 26, No. 1, pp. 50–54, 2011.
- [119] 石崎雅人, 伝康晴. 談話と対話. 東京大学出版会, 2001.
- [120] 河原達也. 音声対話システムの進化と淘汰 -歴史と最近の技術動向-. 人工知能学会誌, Vol. 28, No. 1, pp. 45–51, 2013.
- [121] 中野幹生, 駒谷和範, 船越孝太郎, 中野有紀子, 奥村学. 対話システム. コロナ社, 2015.
- [122] 東中竜一郎, 船越孝太郎. 言語処理学会第 22 会年次大会チュートリアル資料: 対話システムの理論と実践, 2016.
- [123] 東中竜一郎, 船越孝太郎, 荒木雅弘, 塚原裕史, 小林優佳, 水上雅博. Project next nlp 対話タスクにおける雑談対話データの収集と対話破綻アノテーション.
- [124] 内海彰. 言外の意味のコミュニケーション: 語用論概説. 人工知能学会誌, Vol. 18, No. 3, pp. 337–345, 2003.
- [125] 福田拓也, 若林啓. 統計的バックチャネル応答抽出手法. 人工知能学会論文誌, Vol. 33, No. 1, pp. 1–10, 2018.
- [126] 目黒豊美, 東中竜一郎, 堂坂浩二, 南泰浩. 聞き役対話の分析および分析に基づいた対話制御部の構築. 情報処理学会論文誌, Vol. 53, No. 12, pp. 2787–2801, 2012.
- [127] 角森唯子, 大西可奈子, 藤本拓, 角野公亮, 吉村健, 磯田佳徳. カスタマイズ可能なオープンドメイン雑談対話エンジンの開発. 第 84 回言語・音声理解と対話処理研究会, 2018.
- [128] 下岡和也, 徳久良子, 吉村貴克, 星野博之, 渡部生聖. 音声対話ロボットのための傾聴システムの開発. 自然言語処理, Vol. 24, No. 1, pp. 3–74, 2017.

業績リスト

査読付き論文誌

- (1) 光田航, 東中竜一郎, 富田準二, **雑談対話における言外の情報の類型化**, 人工知能学会論文誌, Vol. 31, No. 1, Jan, 2020, pp. 1-10
- (2) 光田航, 東中竜一郎, 青野裕司, **雑談対話における理解結果を伝達する発話の効果**, 知能と情報, 2021 (採録決定)

査読付き国際会議

- (1) Koh Mitsuda, Ryuichiro Higashinaka, and Yoshihiro Matsuo, What information should a dialogue system understand?: Collection and analysis of perceived information in chat-oriented dialogue, Proc. of the 8th International Workshop on Spoken Dialog Systems (IWSDS), June, 2017, pp. 1-9
- (2) Koh Mitsuda, Ryuichiro Higashinaka, and Junji Tomita, Investigating the effect of conveying understanding results in chat-oriented dialogue systems, Proc. of the 8th International Joint Conference on Natural Language Processing (IJCNLP), Nov, 2017, pp. 389-394
- (3) Koh Mitsuda, Ryuichiro Higashinaka, Taichi Katayama, and Junji Tomita, Generating supportive utterances for open-domain argumentative dialogue systems, Proc. of the 10th International Workshop on Spoken Dialog Systems (IWSDS), Apr, 2019, pp. 1-12

報告

- (1) 光田航, 東中竜一郎, 牧野俊朗, 松尾 義博, **雑談対話における言外の情報を推定するためのデータ収集と分析**, 人工知能学会第 30 回全国大会 (JSAI), June, 2016

- (2) 光田航, 東中竜一郎, 松尾義博, **複数の作業者グループを用いた対話における言外の情報の類型化**, 信学技報 NLC2016-31, Dec, 2016
- (3) 光田航, 東中竜一郎, 松尾義博, **対話システムの発話生成における言外の情報の有用性検証**, 人工知能学会第 31 回全国大会 (JSAI), May, 2017
- (4) 光田航, 東中竜一郎, 片山太一, 富田準二, **議論スキームを用いた議論対話における論拠の自動生成**, 人工知能学会第 32 回全国大会 (JSAI), June. 2018