

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

Effectiveness of perceived information on utterance generation in dialogue systems

光田航 東中竜一郎 松尾義博
Koh Mitsuda Ryuichiro Higashinaka Yoshihiro Matsuo

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
NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation

Conveying understanding results of dialogue systems to users with feedback utterances is important. However, in chat-oriented dialogue, feedback utterances are limited to the content of user's utterances. We have collected the information that humans perceive from each utterance (perceived information) in chat-oriented dialogue. Since collected instances can be regarded as system's understanding results and can be used for feedback utterances, we investigated the effectiveness of such feedback utterances for utterance generation. We found that the information other than the user's information that holds only during dialogue can be used as responses. We discuss the reason behind this on the basis of theories in dialogue research. Additionally, we construct a binary classifier which estimates usable utterances created from perceived information.

1. はじめに

対話において、自分の理解を相手に伝えるフィードバック発話は重要である [Litman 04, Raux 06]. 現在の雑談対話システムは、ユーザの発話からキーワードや述語項構造を抽出し、それらを発話生成用のパタンに当てはめ、復唱することで、確認や質問などのフィードバックを行う [Weizenbaum 66, Higashinaka 14]. また、音声認識や発話理解の誤りを抑制するために、ユーザの発話に対して相槌などのフィードバックを行う研究が報告されている [井上 15]. ユーザ発話をそのままフィードバックとして返したり、適切な相槌を行うことで、システムはユーザ発話に適切なフィードバックを返すことができるが、システムが生成しうる発話はユーザ発話の内容に限定されるという問題がある。

我々はこれまでの研究で、雑談対話において、たとえ明示的でなかったとしても、人が発話から一般に理解可能な情報（言外の情報）を収集し、分析を行っている [光田 16b, 光田 16a]. 言外の情報は発話の理解結果であるので、言外の情報を発話に用いることで、ユーザの発話内容に限定されない、理解結果のフィードバック発話が可能であると考えられる。そこで、本研究では、言外の情報をフィードバック発話として利用し、その発話を主観評価することで、発話生成における言外の情報の有用性を検証する。本稿では、収集した言外の情報、および、発話生成の主観評価実験について述べる。さらに、発話生成において低評価となった言外の情報の原因について、考察を行う。また、調査の結果に基づき、言外の情報から生成されうる発話を実際に発話して問題ないかを判定する分類器の構築について述べる。

2. フィードバック発話生成用データの収集

我々が過去に実施した言外の情報の収集、および、そのタイプ分類について述べる。先行研究では、まず、テキストチャットデータを対象に、言外の情報の収集を行った。次に、収集された言外の情報を複数の作業者が人手でグループ化していくことで、言外の情報の分類を作成した。

2.1 言外の情報の収集

我々は以下の手順で言外の情報を収集した [光田 16b]. まず、対話テキストにおける各発話に対して、複数の作業者が言外の情報を記述した。次に、同じ内容を表す言外の情報を人手でまとめ上げた。言外の情報を記述する対象のテキストには、東中らが収集した雑談コーパス [Higashinaka 14] を用いた。このコーパスには、初対面の人同士が、好きな話題についてチャットを行った対話の履歴が収録されている。この中からランダムに選択した 30 対話（作業対象発話数: 1,103 発話）を収集に利用した。その結果、12,723 個の言外の情報が記述され、11,533 個（91%）にまとめ上げられた。

2.2 言外の情報のタイプ分類

収集した言外の情報を利用して、言外の情報のタイプ分類を作成した [光田 16a]. 複数の作業者が相談しながら、類似する言外の情報をグループ化していくことで、揺れの少ないタイプ分類を作成した。作成したタイプ分類を評価するために、3名の作業者が 3000 事例の言外の情報にタイプをアノテーションした結果、Fleiss の κ 係数は 0.69 となり、高い一致率となった。

表 1 に作成された言外の情報のタイプ分類を表す。作業手順から、階層的な分類になっている。レベル 3 が最下層のタイプに対応し、レベル 1 が最上層のタイプに対応している。レベル 3 からレベル 1 に向かってグループ化が行われている。表 1 には言外の情報の分類に加えて、レベル 3 のタイプの概要、および、言外の情報の例が示されている。

3. 言外の情報に基づく発話生成の評価

言外の情報から作成したフィードバック発話の良さを評価するため、既存の対話システムが生成可能な発話と、言外の情報を用いたフィードバック発話の比較を行う。具体的には、比較対象の発話と言外の情報を用いた発話を作成し、それらについて主観評価実験を行った。図 1 に言外の情報を用いた発話の作成と評価の流れを示す。

3.1 発話の作成と主観評価の手順

表 2 に評価対象となる発話を示す。比較対象の発話が 3 種類、言外の情報を用いた発話が 1 種類存在する。これらの 4 種類は以下の通りである。

連絡先: 光田航, 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, 神奈川県横須賀市光の丘 1-1, 239-0847, mitsuda.ko@lab.ntt.co.jp

表 1: 言外の情報のタイプ。A は話者を表し、B は話し相手を表す。丸括弧内の割合は、各タイプが全体のうちに占める割合を示す。

レベル 1	レベル 2	レベル 3	概要	言外の情報の例
思考 (55.4%)	考え (35.8%)	信念 1 (30.7%)	話者の考え、趣向、感情	A は物価が不満だ、A はゲームが好きだ、A は嬉しい
		信念 2 (5.1%)	話し相手に対する話者の信念	A は B に同意した、A は B の話が信じられない
	願望 (19.3%)	願望 1 (9.9%)	話者の主体的な願望	A は富士山に行きたい、A は B に趣味の話がしたい
		願望 2 (9.4%)	話し相手に関連のある話者の願望	A は B に褒められたい、A は B と話を合わせたい
事実 (44.9%)	話者の事実 (37.9%)	属性 (20.2%)	話者のプロフィール、能力、性格	A は女性だ、A は大阪に詳しい、A は正直だ
		行動 (14.4%)	話者の習慣、過去、対話中の様子	A は車を運転する、A は旅行に行った、A は得意気だ
		環境 (3.3%)	話者を含む周囲の人間や生活の環境	A の夫はテレビをよく見る、A の実家は関東にある
	その他の事実 (7.0%)	事実 1 (3.9%)	話者以外に関する確定的な内容	今年は猛暑だ、富士山は紅葉が有名だ
		事実 2 (3.1%)	話者以外に関する意見や予想	農作物が不作になりそうだ、韓国旅行はリーズナブルだ

表 2: 評価対象の発話。発話の例は、“気が向いたら、紅葉を見に富士山に行きます” というユーザ発話に対する応答発話を示す。

評価対象発話	概要	発話の例
対話システム発話	ドコモの雑談対話 API*1 にユーザ発話を入力したときの発話	富士山で働いていますか？
繰り返し発話	ユーザ発話中の述語項構造を繰り返す、人手で作成した発話	富士山に行くんですね
話題応答発話	ユーザ発話中の単語や話題に対して反応する、人手で作成した発話	富士山の紅葉は綺麗そうですね
言外の情報発話	対象の言外の情報を確認する、人手で作成した発話	車を運転するんですね

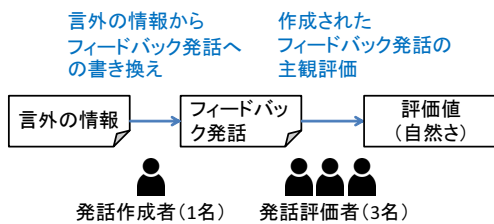


図 1: 言外の情報を用いた発話の作成と評価の流れ

対話システム発話 既存の雑談対話システムとの比較を行うために採用した。

繰り返し発話 既存の対話システムが可能な、述語項構造を用いたフィードバック発話 [Higashinaka 14] との比較を行うために採用した。発話の末尾は必ず「～ですね」で終わる。自動的な作成も可能であるが、解析誤りによる影響をなくすため、ユーザ発話を人手で書き換えることで、発話を作成した。

話題応答発話 理想的な対話システムが返しうる最もよい発話との比較を行うために、キーワードや話題に対して適切に一言で反応する発話を人手で作成したものである。

言外の情報発話 言外の情報を用いたフィードバック発話である。言外の情報から発話への書き換えは人手で行った。発話を作成する際には、内容語の書き換えは行わず、発話の末尾が必ず「～ですね」で終わるようにした。話者と聞き手を表す記号である A と B は、「あなた」と「私」に変え、「あなた」は原則省略するようにした。

発話の作成に用いるユーザ発話、および、言外の情報は表 1 に示した言外情報のタイプに基づき、先行研究で収集した言外の情報の中から選択した。表 1 のレベル 3 のタイプごとに、およそ同数の言外の情報をランダムに抽出し、合わせて 500 個の言外の情報を利用した。比較対象の発話の作成に利用するユーザ発話は、抽出された 500 個の言外の情報のそれぞれについて、それらを記述するために用いられた発話を利用した。抽出されたユーザ発話と言外の情報を利用して、“繰り返し発話”と“言外の情報発話”は 1 名が発話を作成し、“話題応答発話”は 2 名が独立に発話を作成した。“話題応答発話”は、作成された発話の質が発話作成者によって異なる可能性が考

表 3: フィードバック発話の自然さの評価 (7 段階: 1 点～7 点)

評価対象発話	評価者 1	評価者 2	評価者 3	平均
対話システム発話	2.07	2.32	2.20	2.20
繰り返し発話	3.70	4.01	4.56	4.09
話題応答発話 (作成者 1)	4.53	5.38	5.47	5.13
話題応答発話 (作成者 2)	4.47	5.10	5.47	5.01
言外の情報発話	2.72	3.09	3.10	2.97

えられるため、2 名が作成した発話の評価した。以上を合わせて、ユーザ発話 500 発話に対して、計 5 種類 (1 × 3 + 2 × 1)、2500 発話のフィードバック発話の評価に利用した。

作成した各発話が対話システムで利用可能かを評価するために、発話の自然さを評価した。評価者には、フィードバック発話に対して、「システムとユーザとのやりとりは自然かどうか」を、7 段階の評価値 (1 点: とても不自然～7 点: とても自然) として、絶対評価で付与するよう教示を与えた。発話を評価する際には、ユーザ発話とシステムのフィードバック発話と合わせて、ユーザ発話の前の文脈を参照できるようにした。3 名の評価者が独立に評価を実施した。

3.2 発話の自然さの主観評価結果

表 3 に発話の評価した結果を示す。表 3 には、5 種類の発話に対して 3 名の評価者が付与した評価値の平均値が示されている。この表からまず、どの評価者も“話題応答発話 (作成者 1)”，もしくは“話題応答発話 (作成者 2)” に対して最も高い評価値を付与し、“対話システム発話” に対して最も低い評価値を付与したことがわかる。“言外の情報発話” は、“対話システム発話” よりも高い評価値になっているものの、“繰り返し発話” よりも低い結果となった。付与された評価値は、どの評価者も上から“話題応答発話”，“言外の情報発話”，“繰り返し発話”，“話題応答発話 (作成者 1)”，“話題応答発話 (作成者 2)” と並んでおり、作業間で同じ傾向であることがわかる。以降、3 名の平均値を対象に分析を行う。

図 2 に、“言外の情報発話” の評価値を、言外の情報のタイプごとにプロットした図を示す。図 2 には、各タイプの発話に対して、3 名の評価者が付与した評価値の平均が示されている。“言外の情報発話” は、大きく 1～2 点、3 点～4 点、4 点以上の 3 種類に分かれていることがわかる。この 3 種類を低評価タイプ、中評価タイプ、高評価タイプと名付け、各タイプでどのような言外の情報が含まれているかを調査した。説明のために、表 4 に各タイプにおける“言外の情報発話” の例を示す。以降、評価値の低いタイプから順に述べる。

*1 https://dev.smt.docomo.ne.jp/?p=docs.api.page&api_name=dialogue&p_name=api_reference

表 4: 言外の情報発話の評価値の下位 3 発話と上位 3 発話の例。括弧内は評価タイプを示す。評価値は全評価者の平均を示す。

発話	評価	発話	評価	発話	評価
信念 1 (中)	得意気ですね 1.00 ちょっと恥ずかしいですね 1.33 農産物に興味がないですね 1.67 ... 自然観光が好きなのですね 5.33 富士山が好きなのですね 6.00 カッコいい車が好きなのですね 6.33	願望 2 (低)	私に共感してほしいですね 1.00 私と話を続けたいですね 1.00 私に自分を知ってほしいですね 1.00 ... 私と普通の会話がしたいですね 1.67 私をジムに入れたいですね 1.67 私と中国の話をしたいですね 2.00	環境 (中)	あなたの家には PC があるのですね 1.67 主に夫の収入で暮らしているのですね 1.67 あなたの両親は健在なのですね 2.00 ... 親戚は祝い事を大切にしているのですね 5.00 あなたの場所は晴れているのですね 5.33 バスケットチームに入っているのですね 6.00
信念 2 (低)	私に心を開いているのですね 1.00 私に好感を持っているのですね 1.00 私を誤解していたんですね 1.00 ... 私のイメージが変わったのですね 2.00 私の情報に喜んでいるのですね 2.33 私に同感なのですね 2.67	属性 (中)	女性なのですね 1.00 飄々とした人間ですね 1.00 虚弱なんですね 1.00 ... 親切なのですね 5.67 遠出をいとわないのですね 6.33 ワイルドですね 6.33	事実 1 (高)	今は 9 月なのですね 3.00 バイクの費用は人それぞれなのですね 3.33 バイクと自転車は違うのですね 3.67 ... 今年は残暑が厳しいのですね 6.33 曲がったハンドルが特徴なのですね 6.33 ホイールは値段が高いのですね 6.67
願望 1 (低)	話題を映画に変えたいのですね 1.00 夫を自慢したいのですね 1.00 共感したいのですね 1.00 ... 涼しい気候で生活したいのですね 4.33 ドイツに行きたくなったのですね 4.33 移動生活に憧れているのですね 4.67	行動 (中)	どう伝えようか考えているのですね 1.33 次に何を言うか考えているのですね 1.33 話を聞く姿勢を示しているのですね 1.33 ... ハーブティーも飲むのですね 5.00 ナマコの姿に食欲を失うのですね 5.33 この夏よく外へ出かけたのですね 5.33	事実 2 (高)	パソコンは時間を奪うのですね 2.00 お店は有名な割に美味しくないのですね 3.33 筋肉痛は一日遅れでやってくるのですね 3.67 ... 長友選手がチームを引っぱったのですね 6.33 ドイツは安全なのですね 6.33 スポーツバイクは値段が高いのですね 6.67

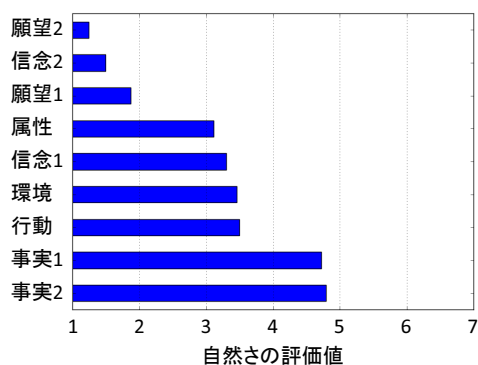


図 2: 各タイプにおける言外の情報発話の評価値

3.2.1 低評価タイプ

低評価タイプは、主に対話中のユーザの思考内容に言及する発話である。特に、ユーザが話し相手（システム）について考えている内容に言及する“願望 2”と“信念 2”の発話に対して、低い評価値が付与されていた。

願望 2 このタイプの発話は、対話に関する、ユーザからシステムへの依頼について言及する発話（例。私に共感してほしいのですね）である。全ての発話に対して、2 点以下の低い評価値が付与されていた。

信念 2 このタイプの発話は、ユーザがシステムに対して抱いた考えについて言及する発話である。例えば、「私に好感を持っているのですね」というように、ポジティブな内容の発話が多く見られたが、低い評価値が付与されていた。

願望 1 このタイプの発話は、対話の進め方に関するユーザ意図に言及する発話（例。話題を変えたいのですね）である。一部、対話を行う最中以外でも成立する内容に言及する発話（例。移動生活に憧れているのですね）も存在し、高い評価値が付与されていた。

3.2.2 中評価タイプ

中評価タイプは、ユーザの性質や振る舞いに言及する発話である。元となった言外の情報には、対話が行われている間でのみ成り立つ場合と、対話の最中以外（対話を行う前や後）でも成り立つ場合が存在する。前者の発話の評価値は低く、後者の発話の評価値は高く判定されていた。また、ネガティブな内容を表す発話（例。頭が固いのですね）には低い評価値が付与されていた。

信念 1 このタイプの発話には、対話中のみ成り立つものとして、対話中のユーザの考えに言及する発話（例。得意気ですね）が存在する。また、対話中以外でも成り立つものとして、ユーザの趣向などに言及する発話（例。富士山が好きなのですね）が存在する。

属性 このタイプの発話は、対話中以外でも成り立つ、ユーザのプロフィールに言及する発話（例。親切なのですね）である。一般的に話題になりやすいもの（出身地、居住地、職業など）には高い評価値が付与され、話題になりにくいもの（年齢、性別、収入、性格など）には低い評価値が付与されていた。

環境 このタイプの発話は、対話中以外でも成り立つユーザの背景の情報に言及する発話である。常識的に考えて自明なもの（例。あなたの家には PC があるのですね）、および、自明ではなく話題になりやすいもの（例。バスケットチームに入っているのですね）が存在する。前者には低い評価値が付与され、後者には高い評価値が付与されていた。

行動 このタイプの発話には、対話中のみ成り立つものとして、対話中のユーザの様子に言及する発話（例。話を聞く姿勢を示しているんですね）が存在する。また、対話中以外も成り立つものとして、対話以前のユーザの習慣や過去に言及する発話（例。この夏よく外に出かけたのですね）が存在する。

3.2.3 高評価タイプ

高評価タイプは、ユーザ以外に関する、一般的な事実に言及する発話である。常識的に考えて自明であり話題になりにくいもの、および、自明ではなく話題になりやすい情報が存在する。前者には低い評価値が付与され、後者には高い評価値が付与されていた。“繰り返し発話”よりも高い評価値が付与された発話が多く、人が実際に行う発話と近い“話題応答発話”と同程度の評価値が付与されていた。

事実 1 このタイプの発話は、常識的に考えて自明であり話題になりにくい内容（例。今は 9 月なのですね）に言及するもの、および、話題になりやすい内容（例。今年は残暑が厳しいのですね）に言及するものが存在する。

事実 2 このタイプの発話は、ポジティブな内容を含む発話（例。ドイツは安全なのですね）、および、ネガティブな内容を含む発話（例。パソコンは時間を奪うのですね）が存在する。後者には低い評価値が付与されていた。

3.3 考察

ユーザ発話から理解可能な言外の情報には、ユーザの思考やプロフィールなど、さまざまな情報が含まれるため、フィードバック発話として適切でない情報も存在する。聞き手に適切な発話を行うためには、協調の原則（会話における当を得た発言のルール）[Grice 75]、および、ポライトネス（会話の参加者の基本的欲求を侵さないための配慮）[Brown 87]に従う必要がある。

主観評価実験の結果は、協調の原理やポライトネスの理論によって説明できると思われる。自明な情報に言及する発話は、協調の原理における量の格率（必要以上の情報を与えてはならない）を違反している。ネガティブな内容や一部のプロフィール情報に言及する発話は、ユーザのポジティブフェイス（承認されたいという欲求）を侵害する。また、対話中のユーザに関する内容に言及する発話は、ユーザに内容を確認したり、その結果を表明する必要が生じさせるため、ネガティブフェイス（行動を制限されたくない）を侵害する。一方、“高評価タイプ”の発話はユーザに直接関係がある内容に言及しておらず、フェイスを侵害していない。

本研究で新たにわかった知見は、対話中にユーザが抱いた考えや対話中のユーザの振る舞いなど、対話中でのみ成り立つ言外の情報は発話すべきでなく、一方で、対話中以外でも成立する願望やユーザ情報は発話に用いてもよく、発話生成に有用であるということである。この知見は、我々が網羅的に収集した言外の情報を発話に用いることで明らかにすることができたと考えている。また、今回の調査から、これまでの理論を確認できたとともに、それぞれがどの程度問題であるかを明らかにすることができたと考えている。

4. 高評価タイプの言外の情報発話の自動分類

フィードバック発話の評価結果から、“言外の情報発話”のうち、高評価タイプの発話は“繰り返し発話”よりもよい評価値になることがわかった。したがって、ユーザ発話に対する“言外の情報発話”を推定できたとして、それが高評価タイプか否かを高精度に判定することができれば、そのままフィードバックとして発話に利用できると考えられる。本節では“言外の情報発話”が高評価タイプか否かを判定する二値分類器を構築し、検証を行う。分類器は一般の発話にも適用できるよう、言外の情報ではなく、主観評価実験で用いた“言外の情報発話”を入力とするように設定した。

分類器の素性として、“言外の情報発話”から、見出し語、日本語語彙大系 [池原 97] の一般名詞意味属性と用言意味属性のカテゴリ番号を抽出し、ユニグラムおよびバイグラムを分類に利用した。訓練データには、主観評価実験に用いた事例とは異なる言外の情報を利用した。具体的には、3,000 事例の言外の情報に高評価タイプ（“事実 1”および“事実 2”）か否かをアノテーションしたデータを利用した。テストデータには、発話評価で利用した 500 事例の“言外の情報発話”を利用した。500 事例の言外の情報のうち、正例は 92 事例（18%）である。学習アルゴリズムにはロジスティック回帰を利用した。

図 3 に“言外の情報発話”の二値分類の評価結果を示す。ロジスティック回帰が出力した正例の確率の高い順に事例を並べた際の PR 曲線が示されている。図 3 から、再現率が 0.3 程度であれば、1.0 に近い適合率で高評価タイプを判定できていることがわかる。再現率が 0.5 程度の場合でも、0.8 と比較的高い適合率で高評価タイプが分類できており、“言外の情報発話”の高評価タイプの判定は高い精度で可能であることがわかる。この結果から、ユーザ発話に対して高評価タイプを含む“

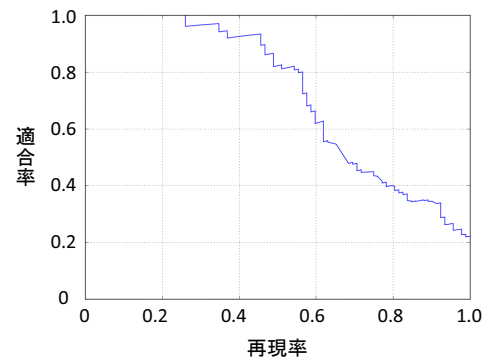


図 3: 言外の情報発話における高評価タイプの自動分類

言外の情報発話”の候補を推定できれば、発話に利用可能な高評価タイプの“言外の情報発話”をフィードバック発話に利用することができることがわかる。

5. おわりに

本稿では、フィードバック発話における言外の情報の有用性を検証した。その結果、対話中のユーザの思考や振る舞い以外のものであれば、フィードバックに利用可能であることがわかった。今後は、ユーザ発話から言外の情報を推定するモデルを作成し、より実用的な評価を行っていききたい。また、タイプごとの傾向を用いて、発話に利用可能な言外の情報をより正確に判定するモデルを構築したい。

参考文献

- [Brown 87] Brown, P. and Levinson, S. C.: *Politeness: Some universals in language usage*, Vol. 4, Cambridge university press (1987)
- [Grice 75] Grice, H. P.: Logic and conversation, in Cole, P. and Morgan, J. eds., *Syntax and semantics*, Vol. 3: Speech acts, pp. 41–58, Academic Press (1975)
- [Higashinaka 14] Higashinaka, R., Imamura, K., Meguro, T., Miyazaki, C., Kobayashi, N., Sugiyama, H., Hirano, T., Makino, T., and Matsuo, Y.: Towards an open domain conversational system fully based on natural language processing, in *Proc. COLING*, pp. 928–939 (2014)
- [Litman 04] Litman, D. J. and Silliman, S.: ITSPOKE: An intelligent tutoring spoken dialogue system, in *Proc. Demonstration papers at HLT-NAACL*, pp. 5–8 (2004)
- [Raux 06] Raux, A., Bohus, D., Langner, B., and Eskenazi, A. W. B. M.: Doing Research on a Deployed Spoken Dialogue System: One Year of Let’s Go! Experience, in *Proc. Interspeech* (2006)
- [Weizenbaum 66] Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine, *Communications of the ACM*, Vol. 9, pp. 36–45 (1966)
- [井上 15] 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, 第 75 回言語・音声理解と対話処理研究会, pp. 21–24 (2015)
- [光田 16a] 光田 航, 東中 竜一郎, 松尾 義博: 複数の作業グループを用いた対話における言外の情報の類型化, 電子情報通信学会技術研究報告 Vol.116 No. 379 言語理解とコミュニケーション, pp. 13–18 (2016)
- [光田 16b] 光田 航, 東中 竜一郎, 牧野 俊郎, 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, pp. 1–4 (2016)
- [池原 97] 池原 悟, 白井 論, 横尾 照男, 中岩 浩巳, 小倉 健太郎, 大山 芳史, 林 良彦: 日本語語彙大系 CD-ROM 版, 岩波書店 (1997)