エージェントが発する「準」自然言語の意味推測：しりとりが与えるヒント

Guessing the meaning of the words expressed by “semi” natural language during “shiritori”

勝 将也[1]　中島 綾乃[1]　菊池 華世[1]　中島 亮一[2]　大澤 正彦[1*]
Masaya Katsu[1], Ayano Nakajima[1], Kayo Kikuchi[1], Ryoichi Nakashima[2], Masahiko Osawa[1]

[1] 日本大学
[1] Nihon University
[2] 京都大学
[2] Kyoto University

Abstract: 本研究では, しりとりをしているという状況が「準」自然言語 (自然言語の各音を「ド」と「ラ」に変換した音) で発せられた単語の推測に与える影響を調べた. 参加者はエージェントが「準」自然言語を発する動画を視聴した後, エージェントが発した音声が意味する単語を回答し, それに対する自信度を報告した. 実験参加者は, しりとりを想定させる・頭文字を知らせる・知らせないの 3 群に分けられた. 結果, しりとりを想定させる群は, 他の 2 群に比べて, 正答率と自信度が有意に高いことが示された. つまり, 「準」自然言語で表現された単語の推測に対して, しりとりをする状況から得られる情報が重要だと考えられる.

1　はじめに

エージェントデザインにおいて, 人間とエージェントのコミュニケーションの継続は重要な課題である. 近年, 対エージェントのコミュニケーションを対人間のコミュニケーションに近づけることを目指し, 自然言語 (人間が生活の中で一般的に使っている言語) 処理の研究が盛んに行われている [1].

ただし, 自然言語を用いたエージェントと人間とのコミュニケーションにおいては, 技術的課題も多い. 対話システムや音声認識などには技術的限界があり, あらゆる状況であらゆる話題の対話を人間と同等に行えるシステムの構築が難しいためである [2][3]. 実際, ボイスユーザインタフェースとのやりとりの中で最も多い障害は, 自然言語処理の失敗によるものであり, 人間に多くのフラストレーションや混乱を引き起こす [4]. 例えば, 駅で経路案内エージェントを使用した実験では, エージェントが案内に 30%以上失敗し, その主な原因は音声認識のミスであった [5]. また, 病院で Pepper を使用した実験では, Pepper の音声認識や顔認識がうまくいかず, 会話が失敗するケースが多く, 対面から 30 分後には全ての患者が Pepper とのコミュニケーションにおいて退屈する様子が観察された [6].

近年, 人間とエージェントのコミュニケーションを継続させるために, 自然言語を発話しないエージェントを用いることが提案されており, ある特定の状況ではそれが有効であることも報告されている [7]. 本研究の目的は, 先行研究で提案された自然言語を発話しないエージェントとのコミュニケーションが成立するために, どのような要素が重要かを明らかにすることである.

2　背景

2.1　適応ギャップ

エージェントとのコミュニケーション継続に関わる要素として, 適応ギャップがある. 小松ら [8] は, エージェントに対して期待する能力の大きさ (F_{before}) と実際のコミュニケーションを通じて感じたエージェントの能力の大きさ (F_{after}) との差 ($F_{after} - F_{before}$) を適応ギャップと定義した. そして, F_{before} と比べて F_{after} が非常に小さいとき, すなわち適応ギャップが大きな負の値となるとき, 人間はエージェントとのコミュニケーションをやめる傾向があると報告した. このことから, 人間とエージェントとのコミュニケーションの継

*連絡先：日本大学文理学部
〒 156-8550 東京都世田谷区桜上水 3-25-40
E-mail: osawa.masahiko@nihon-u.ac.jp

続には, 適応ギャップが正の値であること (あるいは負の値であっても 0 に近いこと) が重要だと考えられる.

負の適応ギャップを回避する一つの方法は, 人間の期待以上のコミュニケーション能力を有するエージェントを構築する（F_{after} を大きくする）ことである. その一例として, 自然言語処理性能の向上がある. ただし, マイクが拾ったノイズ・残響音の除去が不完全であることによる音声入力の問題や, 対複数人との会話における, 話しかけた相手の判定など様々な点において, 解決すべき課題が残されている [2][3]. また, 現在の技術レベルで最大限能力の高いエージェントを構築しても, 人間とエージェントの間でコミュニケーションに失敗することも多い [4][5][6][9]. 能力を高くしたエージェントに対する人間の期待が大きくなってしまうことが一因である. つまり, F_{after} を大きくするのに伴い, F_{before} も大きくなるため, 期待以上の能力を持たせること自体が困難である.

適応ギャップが負の値になることを回避するアプローチとして, 期待される能力の大きさ F_{before} を小さくすることも考えられる. 実際に, 能力を低く見せるため, 非対称な言語コミュニケーション (本論文では, 人間は自然言語を発話し, エージェントは自然言語ではない音で意図伝達を試みるコミュニケーションと定義する) を行うことで, 人間の能力への期待を下げるアプローチが試みられている. 例えば, 小林ら [10] は, システムが相槌だけを返すようなコミュニケーションにおいて, 明滅光源やビープ音でユーザに同じパターンを繰り返し提示するシンプルな相槌表現を行った. 自然言語で「はい」と応答される相槌表現と比較して, 明滅光源やビープ音による応答の方が, 反応が過多ではない, つまりしつこくないと肯定的に評価された. つまり, 明滅光源やビープ音は, 人間とエージェントとの非対称な言語コミュニケーションに効果的である可能性がある.

2.2 「準」自然言語を用いた非対称な言語コミュニケーション

非対称な言語コミュニケーションにおいて, エージェントが自然言語ではない音を発することで自然言語の意味を人間に伝えようとする試みがされている. 人間は文章を理解する際に, それまでの文脈から次に現れる単語を予測している [11][12]. 日常生活でも, 「ただいま」に対する返事は「おかえり」だという予測が行われる場面も多い. そのため, エージェントが自然言語ではない音声を発したとしても, 何らかの手がかりがあれば, それを聞いた人間がその意味を推測し, 自然言語を用いた会話に近いコミュニケーションが可能になるかもしれない.

清丸ら [7] は, ユーザの発話する自然言語を正しく理解できるが, 自然自身は自然言語の単語からアクセントと音韻数を維持して「ド」と「ラ」のみの音で読み上げたもの （これを「と呼ぶ）を発するエージェントを提案し, それを用いた非対称な言語コミュニケーションについて調べた. 彼らの研究では, 状況や文脈, 事前知識をもとにして, 「準」自然言語に対して自然言語的な意味を推測することができると仮説を立てた.

この仮説を検証するために, 清丸ら [7] は人間とエージェントとでしりとりを行わせ, 人間がしりとりを通したコミュニケーションに対してどのような印象を持つかを調査した. このしりとりにおいて, 実験参加者は, 下記の手順でシステムとのやりとりを行うよう制約を設けられた. まず直前にシステムが発した「準」自然言語（例えば最初は「しりとり」に対応する「ドララ」という音声）の意味を推測し, 「(推測した単語・例では「しりとり」) と言ったの？」とシステムに聞き返す. それに対し, システムは, ポジティブな反応を返す. システムと実験参加者がお互いに 1 単語ずつ発話することを 1 往復とし, 実験では 5 往復繰り返すか, 5 分が経過するとしりとりが終了した. その結果, 実験参加者 5 名中 4 名は 5 分以内に 5 往復のしりとりを完了した. また, しりとり終了後に, そのシステムとのコミュニケーションが成立したと思うかを評定した. その結果, 多くの参加者がコミュニケーションができていたと回答し, 自然言語を用いないシステムとも, 非対称な言語コミュニケーションが成立しうることが示唆された. また, この実験では, システムの操作者（実験者）が, しりとりに沿うように単語を選択し, それをシステムに「準」自然言語で発声させていた. しかし, 実験参加者の多くは, コミュニケーションができていたと思ってはいたものの, システムが発した「準」自然言語が示す単語を必ずしも正しく推測できていたわけではなかった. そのため, 「準」自然言語を用いたコミュニケーションでは, 相手が発した単語をわかった気になることによって, ユーザの立場から見ればコミュニケーションが成立するのかもしれない.

2.3 先行研究での課題・本研究の目的

清丸ら [7] の研究では, 人間がシステムが発する「準」自然言語の推測に用いる手がかりを与えるために, しりとりを用いた. また, そのしりとりにおいて, システムが幼稚園児程度の語彙しか持たないことを参加者に教示した. 加えて, エージェントの発した言葉の意味を確認しフィードバックを求めさせた. これらの様々な工夫や教示の中には, 「準」自然言語の自然言語的な意味を推測することに影響した要因もあれば, 影響しなかった要因もあると考えられる. そこで本研究では, しりとりをすることに着目し, それが「準」自然言語として発

せられた単語を正しく推測できているという自信に及ぼす影響を検討する.

具体的には, しりとりを行うことで, 何らかの手がかりが与えられ, それが「準」自然言語の発話を推測することに有効であるかを検討する. そこで, しりとりをするという状況を設定することで, 実験参加者が「準」自然言語で発話された単語の推測ができるようになるかを調べる. また, しりとりでは自分が先に単語を発すると相手が発する単語の頭文字がわかる. それにより相手が発する単語の候補を絞り込める. そのため, 頭文字を知っていることが, その単語を正しく推測できているという自信を持ちやすくなる要因かもしれない. そのために, 実験参加者を, ロボットとしりとりを行うことを想定させる群（以後, しりとり想定群）, あらかじめ発せられる音声の単語の頭文字を知らせる群（以後, 頭文字既知群）, ロボットの発する単語についての情報を与えない群（以後, 無情報群）に分け, それぞれの群における単語推測に対する回答を比較した. 「準」自然言語の発話の推測に対し, 手がかりを与えることが有効かを検証するために, 無情報群を設定した. もし手がかりが有効であるならば, しりとり想定群と頭文字既知群の自信度は, 無情報群よりも高くなると予想される.

3 方法

3.1 参加者

20 歳以上の日本語を母語とする男女をクラウドソーシングサービス[1]で募集し, 356 名が Web アンケートを用いた実験に参加した. 本実験は日本大学文理学部研究倫理委員会の承認を得て実施した. 全員が実験内容に同意したうえでアンケートへの回答を行った.

3.2 刺激

アンケートは web アンケート作成システム[2]を用いて作成された. 実験参加者がエージェントが「準」自然言語を発する動画を視聴し, その後その動画で発せられた音声に関する質問に回答した.

実験で使用した動画は, 映像として表示する静止画 (図 1, [13]) と音声データを組み合わせることで作成した. 音声データは以下の手順で作成した. まず, 清丸ら [7] が用いた音声のもととなった単語の中から, アクセントと音韻数の異なる 12 種類の自然言語の単語を選んだ (表 1). それぞれの単語に対し, アクセントと音韻数を維持したまま各音を「ド」あるいは「ラ」の音に置き換えて, ロボットが発する音声とした. 画像と音声を

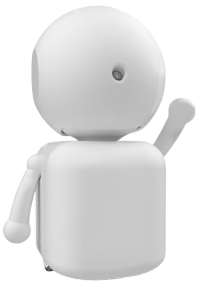


図 1: 「準」自然言語エージェント. 動画に表示する画像として使用した.

表 1: 動画で再生される音として使用した単語の単語親密度

単語	単語親密度
あめ	6.469
かも	5.781
くに	6.406
もり	5.688
ねずみ	6.281
ひよこ	6.156
まんが	6.406
らくだ	5.875
うめぼし	6.125
きつつき	5.875
しまうま	6.094
みそしる	6.406

組み合わせる際, 動画中音声が発せられる前に 1 秒間, 後に適切な空白時間 (3 秒 − 音声時間 [秒]) を設け, 各動画の再生時間を 4 秒間に統一した.

実験後に, 選んだ単語について音声に関する単語親密度 [14] を調べたところ, すべて 7 段階中 5.6 以上であった（表 1）. 天野ら [14] に記載されている 69,084 個の単語における単語親密度の平均値が約 4.3, 標準偏差が約 1.3 であり, これらは比較的親密度の高い単語だったと考えられる.

3.3 手続き

各実験参加者は, 自分の持つ PC で web アンケートページにアクセスし実験に参加した. その際, 彼らは, しりとり想定群, 頭文字既知群, 無情報群にランダムに振り分けられた. 各群の参加者は, アンケートの前に, 動画の音声が問題なく聞こえることを確認するために, 確認用動画で再生される音声 (自然言語で「かくれんぼ」という単語を提示) を聞き, それを入力するように求められた.

[1]「CrowdWorks」: https://crowdworks.jp/
[2]「SoSciSurvey」: https://www.soscisurvey.de/

各群の参加者は, 動画の再生ボタンを自分でクリックしてロボットが「準」自然言語を発する動画を視聴した. その際, しりとり想定群では自分がしりとりで発した単語が画面上に表示された（例：このロボットとしりとりをしている状況を想定してください. そのやり取りの途中で, あなたは「たぬき」と言いました）. 頭文字既知群ではロボットが発する単語の頭文字が画面上に表示された（例：動画内でロボットは「き」から始まる単語を言います）. これらの群の参加者は, その情報を確認したうえで, 動画を再生した. 無情報群の参加者は, 何の情報も与えられない状態で動画を再生した. 動画を一度再生すると動画が非表示となり, 各群の参加者は「次へ」のボタンをクリックし回答画面に進んだ.

回答画面では, 動画で発せられた音声は何の単語を意味していたかを推測し, それをひらがなで回答欄に入力した. その際, 「準」自然言語として発せられた音声は, 広辞苑に含まれる名詞を意味しており, 固有名詞ではないことと, また実験中に広辞苑に含まれる単語かどうか調べる必要はないことが教示された. 単語の回答後, その回答に対する自信度を 7 段階（1: 全く自信がない〜7: とても自信がある）で評定した. これを 12 種類の動画に対してランダムな順で行った.

3.4 データ分析

表 2 に示す基準をすべて満たす参加者を, 実験の教示に従って回答した実験参加者として, データ分析の対象とした. その結果, しりとり想定群 116 名（39.2±10.2 歳, 男性 45 名, 女性 71 名）, 頭文字既知群 89 名（41.9±10.4 歳, 男性 35 名, 女性 53 名）, 無情報群 84 名（42.0±10.4 歳, 男性 37 名, 女性 47 名）が分析対象となった.

各群の回答の正答率 (参加者の回答した単語と, 「準」自然言語のもとになった単語が合致した場合を正答と定義した), その回答に対する自信度評定値を比較した. 正答率, 自信度評定値に対して, それぞれ一要因分散分析を行った.

4 結果

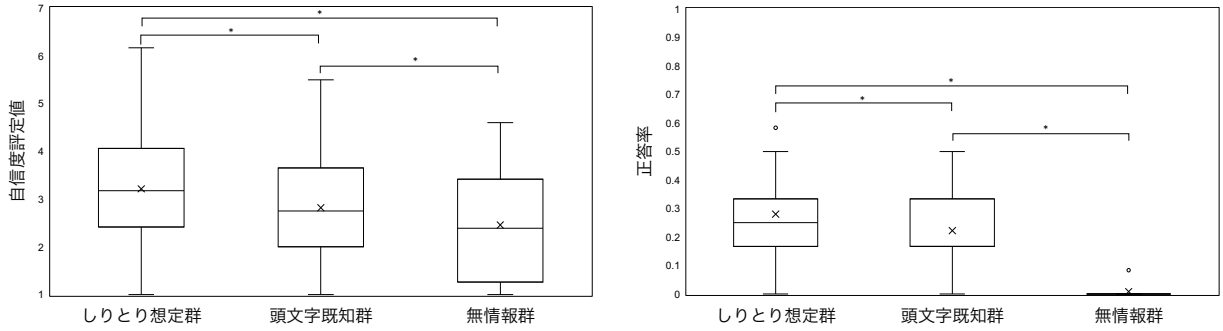
各群における自信度評定値と正答率を図 2 に示す. 自信度評定値において, 群の主効果が有意であった ($F(2, 286) = 11.33$, $p < 0.001$, $\eta^2 = 0.073$). Holm 法による多重比較の結果, しりとり想定群, 頭文字既知群, 無情報群の順に自信度評定値が高かった (すべての水準間で ps < 0.035). しりとりは, 「準」自然言語で発話された単語を「わかった気にさせる」影響があると考えられる. エージェントが発する「準」自然言語が示す単語について, 頭文字を知っていると「準」自然言語が意味する単語の候補が大幅に限定されるため, 「準」自然言語が意味する単語についての自分自身の推測が正しいという気になりやすいと考えられる. さらに, しりとり想定群では, 頭文字既知群よりも自信度評定値が高いため, しりとりにおける頭文字を知っていること以外の他の要因による影響も考えられる. 例えば, しりとりをしていて自分がある単語を言ったという認識が重要である可能性がある. つまり, 自身が発話した単語と, しりとりでよく現れる典型的な単語対のパターン (例：「リンゴ」と言われたら次は「ゴリラ」) に基づき, 次に発話される単語を推測したのかもしれない.

表 2: 分析対象とする参加者を選ぶ基準. 全てを満たす参加者のデータを分析対象とする.

基準 1	年齢の質問に対して半角で 20 以上の数字を入力している
基準 2	動画再生確認の質問に対して正しく回答している
基準 3	動画を一度だけ再生している
基準 4	ひらがなで回答している
基準 5	指定した頭文字の単語を 9 回以上回答している (無情報群を除く)
基準 6	広辞苑に含まれる名詞で 8 回以上回答している

また, 正答率においても, 群の主効果が有意であった ($F(2, 286) = 183.07$, $p < .001$, $\eta^2 = 0.561$). 多重比較の結果, しりとり想定群, 頭文字既知群, 無情報群の順に正答率が高かった (ps < 0.001). しりとりでよく現れる典型的な単語対のパターンに基づき, 次に発話される単語を推測し回答したことで, 自信度の向上と同様にしりとり想定群が頭文字既知群より正答率が高くなったのかもしれない. また, 無情報群では, 単語に関する手がかりとなる情報が無いため推測自体が難しく, 76 名は 12 問全て不正解だったが, 8 名は 12 問中 1 問正解だった. 本実験で用いた単語全ての単語親密度が高かったことから, アクセントと音韻数が一致している単語の中でも, 一部の参加者にとって推測しやすい単語が含まれていたのかもしれない.

しりとり想定群, 頭文字既知群ともに正答率 50%以上であった「らくだ」, 「ねずみ」は, しりとり想定群では, それぞれ自信度評定値が 3.48, 3.21 であり, 頭文字既知群では, 2.94, 2.64 であった. 一方, 正答率が 5%以下であった「かも」, 「くに」は, しりとり想定群では, それぞれ自信度評定値が 3.79, 3.72 であり, 頭文字既知群では, 3.19, 3.15 であった. つまり, しりとり想定群, 頭文字既知群に関係なく, 正答率が高い単語を必ずしも自信を持って回答しているわけではない.



(a) 自信度評定値　　**(b)** 正答率

図 2: 実験結果. 図における箱は四分位範囲, ひげの上端と下端は外れ値を除く最大値と最小値を示す. また, ドットは外れ値, ×印は平均値, *は $p<0.05$ を示す.

5　考察

5.1　本研究の有効性

1 章で述べたように, 人間とエージェントが自然言語を用いた会話を試みる場合, 現状の技術レベルで最大限能力の高いエージェントを構築しても, エージェントの振る舞いがその期待と乖離し, コミュニケーションが失敗するケースが多い. 清丸ら [7] は, 自然言語を発話しないことで, エージェントに対する期待を下げつつ,「準」自然言語を発し, 自然言語の意味を人間にわかった気にさせるコミュニケーション手法を提案したが, コミュニケーションが成立する理由については, 十分に検討されていなかった. 本研究では, コミュニケーションの成立を相手の発話した単語を推測できることと定義し, 検討を行った.

本研究において, しりとり想定群, 頭文字既知群, 無情報群の順に自信度評定値が高かったことから, しりとりをしている状態の, 頭文字を知っているという要素や, それ以外の要素の影響で,「準」自然言語を正しく推測できているという自信を持ちやすくなったと言える. また, しりとり想定群, 頭文字既知群に関係なく, 正答率が高い単語を必ずしも自信を持って回答しているわけではなかった. これらの結果より, 何らかのヒントがある状態で, エージェントが「準」自然言語を発話することで, 人間に対して, 単語の意味を「わかった気にさせる」可能性があると考えられる. 自分自身が推測した単語に対する自信を持ちやすくなると, エージェントの発話内容にミスがあった場合でも, 人間がエージェントの発話内容を推測し, エージェントとのコミュニケーションをとれるようになるかもしれない. つまりエージェントの発話内容のミスによって, 人間のエージェントに対する期待が下がる機会を減らすことができる可能性がある. 本研究は, 人間とエージェントの非対称な言語コミュニケーションをうまく成立させるための, 一つのヒントとなる知見を提供していると考えている.

5.2　本研究の限界

実験において, しりとり想定群が, 頭文字既知群よりも自信度評定値, 正答率が高かったことから,「準」自然言語を聞く前に頭文字の情報を得ること以外に, しりとり特有の要因による影響があると考えられる. しりとり特有の要因として, 自身が発話した単語と, しりとりにおける典型的な単語対のパターン (例：リンゴ, ゴリラなど) に基づき, 次に発話される単語を予測することが考えられる. しかしながら, 典型的な単語対のパターンが多くの人間で一致しているのか, 個人間に大きなばらつきがあるかはわからない. しりとりにおける典型的な単語対のパターンによって,「準」自然言語音声が表す単語の推測への自信度が変わるかを調べるためには, しりとりにおける単語対のパターンを詳細に調べる必要があるだろう.

本研究のしりとり想定群の参加者は, 動画教示前に, 自身が発話した単語が表示され, しりとりを想定するように教示された. 一方で, 清丸ら [7] の実験では, エージェントと対面し, 実際にしりとりによる非対称な言語コミュニケーションをしていた. また, 本研究では,「準」自然言語が示す単語の推測に焦点を当てたため, しりとりというコミュニケーションの一部についてしか検討できていない. しりとりでは, 相手が言った単語の理解だけでなく, それに対して自分自身の単語の発話も重要である. 実験参加者がエージェントと対面し, 実際にしりとりを行いながら自身の考えで単語を選定し, 一回のやりとりごとに自信度を調査することで, しりとりを繰り返すことによるコミュニケーションの時系列的な変化を明らかにできると考えられる.

本研究では「準」自然言語を発するエージェントと

の非対称な言語コミュニケーションの成立に重要な要素を検討するために, しりとりをするという要素に着目し調査した. 一方, 清丸ら [7] の実験では, それ以外にもさまざまな教示やエージェントの振る舞いに関する工夫がされていた. そのため, 他の教示や工夫もあわせて詳細に検討することで, エージェントデザインやコミュニケーション方法を設計する必要がある.

6 まとめ

本研究では, 「準」自然言語を発するエージェントとのコミュニケーションに関連して, しりとりをすることが有効かを調べた. その結果, しりとりをしている状況から得られる手がかりによって, 「準」自然言語を正しく推測できているという自信を持ちやすくなったことが示唆された. つまり, しりとりは「準」自然言語を用いた非対称なコミュニケーションをスムーズにするのに有効だと考えられる.

謝辞

本研究は, 孫正義育英財団の助成を受けた.

参考文献

[1] Young, T., Hazarika, D., & Poria, S.: Recent trends in deep learning based natural language processing. IEEE Computational Intelligence Magazine, 13(3), 55–75, (2018).

[2] Honig, S., & Oron-Gilad, T.: Understanding and resolving failures in human-robot interaction: Literature review and model development. Frontiers in Psychology, 9, 861, (2018).

[3] Skantze, G.: Turn-taking in conversational systems and human-robot interaction: a review. Computer Speech & Language, 67, 101178, (2021).

[4] Myers, C., Furqan, A., Nebolsky, J., Caro, K., & Zhu, J.: Patterns for how users overcome obstacles in voice user interfaces. In Proceedings of the 2018 CHI conference on human factors in computing systems, 1–7, (2013).

[5] Shiomi, M., Sakamoto, D., Kanda, T., Ishi, C. T., Ishiguro, H., & Hagita, N.: A semi-autonomous commu- nication robot—a field trial at a train station. In 2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), 303–310, (2008).

[6] 野口 博史.: コミュニケーションロボットの入院中高齢者への活用事例 システム/制御/情報 (システム制御情報学会誌), 66(2), 45–49, (2022).

[7] 清丸 寛一, 大澤 正彦, 今井 倫太.: 予測的認知を用いた非自然言語による言語的コミュニケーション 第 6 回汎用人工知能研究会, (2017).

[8] 小松 孝徳, 山田 誠二.: 適応ギャップがユーザのエージェントに対する印象変化に与える影響 人工知能学会論文誌, 24(2), 232–240, (2009).

[9] 東中 竜一郎, 荒木 雅弘, 塚原 裕史, 水上 雅博.: 雑談対話システムにおける対話破綻を生じさせる発話の類型化 自然言語処理, 29(2), 443–466, (2022).

[10] 小林 一樹, 船越 孝太郎, 小松 孝徳, 山田 誠二, 中野 幹生.: ASE に基づく相槌によるロボットとの対話体験の向上 人工知能学会論文誌, 30(4), 604–612, (2015).

[11] Altmann, G. T., & Kamide, Y.: Incremental interpretation at verbs: Restricting the domain of subsequent reference. Cognition, 73(3), 247-264, (1999).

[12] Kleiman, G. M. Sentence frame contexts and lexical decisions: Sentence-acceptability and word-relatedness effects. Memory & Cognition, 8(4), 336–344, (1980).

[13] 福田 聡子, 澤田 志織, 川崎 邦将, 奥岡 耕平, 大澤 正彦, 長田 茂美, 今井 倫太.: 適応ギャップ理論を拡張したインタラクションデザインの提案 HAI シンポジウム, (2018).

[14] 天野 成昭, 近藤 公久.: NTT データベースシリーズ 日本語の語彙特性第 1 巻 単語親密度 三省堂, (1999).