人工知能学会研究会資料
SIG-AGI-006-03

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

Predictive Recognition-based Linguistic Communication Without Using Natural Language

清丸 寛一 [1*]　大澤 正彦 [2, 3]　今井 倫太 [2]
Hirokazu Kiyomaru[1]　Masahiko Osawa[2, 3]　Michita Imai[2]

[1] 京都大学大学院情報学研究科
[1] Graduate School of Informatics, Kyoto University
[2] 慶應義塾大学大学院理工学研究科
[2] Keio University Graduate School of Science and Technology
[3] 日本学術振興会特別研究員
[3] Research Fellow of the Japan Society for the Promotion of Science

Abstract:
We communicate naturally by predictively recognizing utterance based on situation, context, and knowledges. We think that predictive recognition makes robots without ability to speak natural language possible to do linguistic communication. In this paper, we focus on Shiritori, a word game, as a simple linguistic communication to verify the hypothesis. Shiritori restricts players' utterance because of the rules, and also has several typical reply patterns. Therefore, we can easily predict partner's utterances compared with normal conversations. In order to reply predictable words, we construct kindergartener-level vocabulary. Furthermore, we make audio data with the cooperation of voice actors for the rich expression of the words. Experimental result implied that we can play Shiritori even if a player does not speak natural language.

1　はじめに

人間は状況や文脈，事前知識をもとに，相手の発話に対して予測的な認知を行いながらコミュニケーションしている．例えば，人間はペットの動物に対して自然言語で話しかけ，ペットの非言語的な振る舞いを解釈してコミュニケーションを行うことがある．このコミュニケーションが成り立つのは,「ただいま」に対する返事は「おかえり」だろうといった応答に対する予測，すなわち予測的な認知が応答の解釈に先行して行われ，応答が予測の範囲内であるならば，ペットの本来の意図とは関係なく，予測の内容を振る舞いの解釈とするからだと考えられる．

著者らは，予測的な認知を利用すれば，自然言語を話せないロボットも言語的なコミュニケーションを行えるという仮説を立てている．しかしながら，一般のコミュニケーションは応答の自由度が高いため，予測的な認知を引き出すための要素や条件を検証するための例題としては不適当である．そこで本稿では，簡単な言語的なコミュニケーションとして，しりとりを例題に仮説検証を行う．しりとりには，発話の最初の文字の制約や典型的な返答パターンが存在する．したがって，一般のコミュニケーションと比べて相手の次の発話が予測が容易であり，コミュニケーションが成立するための条件を検証しやすい．

次章以降の構成は以下の通りである．第 2 章では，関連研究についてまとめる．第 3 章では，予測的な認知を引き出すためのシステムの構築方法について述べ，第 4 章で評価実験の結果を考察したのち，第 5 章をまとめとする．

*連絡先：京都大学大学院情報学研究科
〒 606-8501 京都府京都市左京区吉田本町
E-mail: kiyomaru@nlp.ist.i.kyoto-u.ac.jp

2　関連研究

2.1　Artificial Subtle Expression

表情や視線，仕草などの非言語情報，また声の大きさや高さなどの情報は subtle expressions と呼ばれ，人間同士のコミュニケーションにおいて重要な役割を果たすことが指摘されている．Artificial Subtle Expression (ASE) は，人工物に適した単純な情報を表出すること

で，subtle expressions と同様に，内部状態を直感的に伝達するものである [1]．船越らは，音声対話エージェントとユーザ間の対話を円滑化することを目的として，明滅光源を用いて内部状態を表出するエージェントを用いてユーザとしりとりを行い，発話の衝突が軽減すること，エージェントに対してユーザが抱く印象が向上することを報告している [2]．

ASE は，エージェントの内部状態を単純な刺激として提示する機構であり，刺激を観測した人間はエージェントの内部状態をボトムアップ的に解釈する．本研究で扱うシステムは，非自然言語の情報を人間に提示する点で ASE と共通しているが，人間が状況や文脈を踏まえてシステムの応答をトップダウン的に解釈する必要がある点で異なる．したがって本研究は，状況に応じて同一あるいは類似した表出に異なる意味を与えられる，ASE の発展として位置づけられる．

2.2 Predictive Coding

Predictive coding は，脳は常に次に受け取る入力について予測を行っているという脳情報処理に関する理論である [3]．Predictive coding によれば，脳は学習で獲得した内部予測モデルに従って上の層から下の層へトップダウンの予測信号を伝達しており，予測信号と入力刺激の誤差情報が下の層から上の層へと伝達され，内部予測モデルが更新される．内部予測モデルの学習の経過に伴い，外界からのボトムアップ入力に対して内部予測モデルからのトップダウン入力の影響が強くなることが報告されている．

本研究で扱うしりとりには最初の文字の制約と典型的な応答パターンが存在するため，トップダウン入力の働きによって発話の内容が解釈できると考えられる．

3 予測的な認知を引き出す発話

本章では，予測的認知を引き出すための語彙の整備，音声データの整備，および応答の生成について説明する．

3.1 語彙の整備

予測的な認知は，可能性のある応答の数が少ないほど的確に働くと考えられる．しりとりは，最初の文字に対する制約はあるものの，返答として可能な語の数は依然として膨大であり，システムの返答に関する仮定がない状況で適切な予測をすることは困難である．そこで本研究では，幼稚園児程度の語彙を整備することで，難しい語が返答として得られる可能性を排除し，予測的な認知を正しく働かせることを目指す．

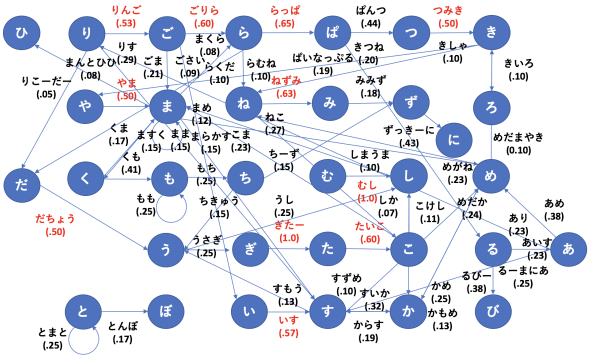

図 1: システムの語彙とその生起確率のグラフ．生起確率は，低頻度語を含めて計算している．赤字は遷移する確率が 50%を越えていることを表す．左下に孤立したグラフは，低頻度の語を除外したことによるものである．

幼稚園児程度の語彙を収集するために，神奈川県内の幼稚園の協力のもと，幼稚園児がしりとりを行う際に用いる語彙を収集した．合計 15 名の幼稚園児に幼稚園教諭とそれぞれ 2 回ずつしりとりを行ってもらった結果，451 回の発話，230 種類の語彙がしりとりのコミュニケーションの中から得られた．登場した語のうち，一度しか現れなかった語は 149 種類であった．

全ての語彙のうち，頻度が 1 回の語と語尾が「ん」の語を除外したものを提案システムの語彙とする．図 1 にシステムの語彙を示す．図 1 より，しりとりには典型的な応答パターンが存在し，語の出現頻度に偏りがあることが分かる．

3.2 音声データの整備

語のニュアンスを適切に伝えることで，的確な予測的な認知を促すことができると考えられる．既存の音声合成ソフトウェアの中には人間の声に近い自然な音質を提供しているものもあるが，それらは自然言語を発話することを目的に設計されており，限られた音で様々なニュアンスを伝えることに最適化されたものは存在しない．

そこで本研究では，芸能プロダクション Artist Crew の協力のもと，収集した語の読み上げをプロの声優に依頼し，音声データを作成した．語の読み上げは，それぞれ出現回数分だけ行った．さらに，しりとり以外のコミュニケーションのために，喜び，悲しみ，そして怒りの感情についてそれぞれ 5 回ずつ音声データを収録した．感情の音声データは，著者らが実際に聞き，喜び，悲しみ，怒り，不明のうちどれに該当するかラベル付けを行い，不明以外でラベルが一致した音声のみを使用する．

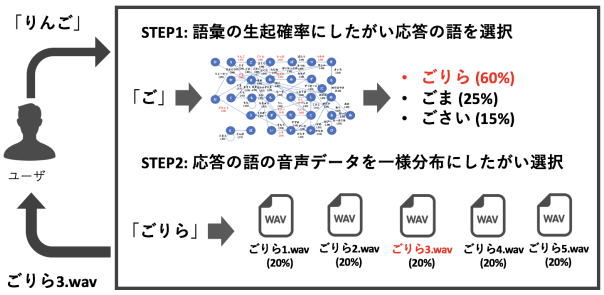

図 2: システムの処理の流れ．ユーザの発話は実験実施者が入力する．

3.3 応答の生成

処理の流れを図 2 に示す．ユーザの発話は，実験実施者がシステムに入力する．その入力をもとに，語の生起確率にしたがい，応答の発話の内容を決定する．さらに，発話の内容を表す音声が複数ある場合，その中の一つを無作為に選択し出力とする．

4 実験

非自然言語によって言語的なコミュニケーションが可能であるか検証するために評価実験を行う．

4.1 人間同士のコミュニケーション

予備実験として，人間同士のコミュニケーションにおいて，一方が自然言語を用いなくてもしりとりが可能であるか検証する．実験は芸能プロダクション Artist crew 所属の 2 人の声優の協力のもと行った．

4.1.1 実験条件

実験では，片方が自然言語を発話し，他方が自然言語を発話しない状況でしりとりを行う．さらに，自然言語を発話しない側の振る舞いについて，二つの設定で実験を行う．一つ目の設定では，表現した内容が相手の解釈と異なれば，ネガティブなフィードバックを与える．この設定で，互いの発話の内容を正しく理解しながらコミュニケーションを進めることが可能であるか検証する．二つ目の設定では，表現した内容が相手の解釈が異なる場合でも，ポジティブなフィードバックを与える．この設定で，予測的な認知によって得られる解釈を全て許容したとき，コミュニケーションがどのように変化するか検証する．

4.1.2 実験結果

一つ目の設定では，互いに全ての発話の内容を正しく理解することを課し，しりとりが 5 往復するまでに 3 分 8 秒を要した．その一方，二つ目の設定では，予測を全て許容した結果，2 分 52 秒でしりとりが 5 往復した．

実験終了後，実験に参加した声優らにインタビューを行ったところ，非自然言語の発話および解釈という不慣れな状況，さらに実験の様子を録音・録画されている状況に緊張し，平常時であれば容易に想像がつくはずの語が思い浮かばなかったことを指摘された．このことは，緊張が予測的な認知を妨げる可能性があることを示唆している．

4.2 人間とシステムのコミュニケーション

構築したシステムを用いて，人間と自然言語を話せないシステムがしりとりをできるか検証する．実験は，10 代，20 代の男女 5 名の被験者の協力のもとで行った．

4.2.1 実験条件

実験は，システムが最初に「しりとり」に対応する非自然言語を発話して始まり，しりとりが 5 往復以上続く，あるいは 5 分が経過した時点で終了とする．被験者には，実験の注意点として以下の三点を説明する．

- システムはしりとり以外の通常のコミュニケーションも可能である．
- 被験者もまた，自由にコミュニケーションを取ることが可能である．しりとりの返答以外の質問や話しかけを行っても良い．必要がなければ，行わなくても良い．
- システムは幼稚園児程度の語彙しか持たない．ある文字から始まる語を持たない，または全て使い切ってしまった場合，答えられる語がないことを非自然言語で伝える．

システムは被験者の解釈が表現した内容と異なる場合でも，ポジティブなフィードバックを与える．実験終了後に，どの程度コミュニケーションが成立したと感じたか 7 段階評価でアンケートを行う．

4.2.2 実験結果

被験者がしりとりの 5 往復に要した時間と，アンケートの結果を表 1 に示す．システムとのコミュニケーションが円滑に行えた被験者は，システムに問いかけを行

い，フィードバックの発話がポジティブであるかネガティブであるか適切に解釈できていた．コミュニケーションが円滑に行えなかった被験者は，システムの発話のうち，しりとりの返答としての発話とその他のフィードバックとしての発話を混同し，ターン・テーキングが不明確になる傾向があった．

表 2 に，アンケートにおいてコミュニケーションが完全に成立していたと答えた被験者の実験結果を示す．表 2 より，被験者がシステムと行った 9 回のやり取りのうち 4 回で被験者の発話の解釈とシステムの実際の発話に相違があることが分かる．コミュニケーションのおよそ半分が誤った解釈にしたがって進行している一方で，コミュニケーションが完全に成立していると感じられたことは非常に興味深い．これは，適切なフィードバックを与えれば，音声認識や言語理解に失敗したとしても，予測的な認知のはたらきを利用することで，ユーザの立場から見れば円滑なコミュニケーションが実現できる可能性があることを示唆している．

表 1: 被験者がしりとりの 5 往復に要した時間とアンケートの結果．アンケート結果の数字はコミュニケーションについて「1: 完全に成立していなかった」から「7: 完全に成立していた」の七段階評価を表している．×は，しりとりが 5 往復する前に 5 分が経過したことを示している．

	5 往復に要した時間	アンケート
23 歳男性	3 分 16 秒	6
29 歳女性	1 分 41 秒	3
18 歳女性	1 分 10 秒	7
24 歳男性	1 分 30 秒	5
24 歳男性	×	6

5 おわりに

本稿では，しりとりを例題として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の実現可能性を検討した．しりとりのコミュニケーションにおいて予測的な認知を引き出すために，幼稚園児が行ったしりとりのログデータから，しりとりにおいて利用されやすい幼稚園児程度の語彙を整備した．さらに，語のニュアンスを限られた音で伝えるために，プロの声優の協力のもと，音声データを作成した．

実験から，対話においてしりとりの返答としての発話と，それ以外の発話の違いが区別できれば，適切なターン・テーキングができ，自然言語を使わず言語的なコミュニケーションが行える場合があることが分かった．今後，発話の種類が識別できるための要素を検証するために，語彙の仮定を教示しない設定，出力の音声データを無作為に選択する設定で実験を行う．また，ロボットの身体的な表現が予測的な認知に与える影響についても検証する．

表 2: アンケートでコミュニケーションが完全に成立していたと回答した被験者がシステムとしりとりを行った際の被験者の発話，エージェントの発話，および被験者が認識した発話．×は応答可能な発話がなかったことを示している．

被験者	システム	ユーザの認識
りんご	ごりら	ごりら
らっぱ	**ぱんつ**	**ぱんや**
やさい	いす	いす
すいか	**かめ**	**かに**
にじ	×	×
にんき	**きしゃ**	**きいろ**
ろうそく	**くも**	**くま**
まる	るーまにあ	るーまにあ
あにめ	め	め

謝辞

本研究は，全脳アーキテクチャ若手の会と一般財団法人 ZERO 財団の助成を受けた．また，幼稚園児の語彙の収集に際しては神奈川県内の幼稚園に協力を頂いたほか，音声データの作成とロールプレイングに際しては株式会社 Artist crew の皆様に協力を頂いた．

参考文献

[1] 小松孝徳, 山田誠二, 小林一樹, 船越孝太郎, 中野幹生: Artificial Subtle Expressions: エージェントの内部状態を直感的に伝達する手法の提案, 人工知能学会誌, Vol. 25, No. 6, pp. 773–741 (2010)

[2] 船越孝太郎, 小林一樹, 中野幹生, 山田誠二, 北村泰彦, 辻野広司: Artificial Subtle Expression としての明滅光源による音声対話の円滑化, 電子情報通信学会論文誌 *A*, Vol. J92-A, No. 11, pp. 818–827 (2009)

[3] Rajesh P. N. Rao, Dana H. Ballard: Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects, *Nature Neuroscience* , Vol. 2, No. 1, pp. 79–87 (1999)