原著論文

人-人，人-ヒューマンエージェントの 社会的インタラクションにおける 言語・非言語行為の配置規則†

榎本 美香 *1・中野 有紀子 *2

本研究では、エージェントを相手にした時のインタラクション方略に人相手のときとの異同があることを示し、人間の行動モデルをベースにヒューマンエージェントを実装するとき、考慮すべき人間のインタラクション方略のあることを提案する。ここでは、パソコン操作課題における人と人、人とエージェントの対話を素材として、どのように言語・非言語行為がインタラクションの中で配置されるのかを分析することで、エージェントに対したときに選択される方略を明らかにする。まず、人対人と人対エージェント対話の基礎的特徴を観察し、人対エージェントの対話では人の発話量が少なく、相づちや応答が稀にしか差し挟まれないことを示す。次に、非言語行為を含めた人対人の行為の配置規則を定式化し、人対エージェントのインタラクションにおいてこの規則がどのように破られるかを示す。そして、この違反が、相づちや応答の変わりに、相手発話への理解を示すためになされた補償的行為であることを明らかにする。

キーワード：ヒューマンエージェント，社会的インタラクション，非言語行為，言語／非言語行為の交替規則，話者交替規則

1. はじめに

本研究では、社会的インタラクションを「何らかの社会的規範に基づいた参加者たちの行為のやりとり」と定義する。そして、人と類似する身体を持ったエージェントと対話するとき、人はどのように発話や身体の動きをインタラクションの中で配置するのかを明らかにする。

本研究の動機は、システムの現状とユーザの解決すべき問題を把握し、ユーザとのインタラクションを通じて共に問題解決を図るヘルプエージェントシステムを実現することにある。このエージェントは、言語情報によって情報を提示するだけでなく、ジェスチャなどの非言語情報を駆使することで、ユーザに最も効率的な問題解決を授けるものである。我々はこれまで、パソコン操作課題時の対話における人間行動の基礎的知見を得てきた。榎本ほか(2006)では、マルチモダリティを用いた社会的インタラクションにおける人間の行動パターンを明らかにし、中野ほか(2007)やMurata

et al.(2007)、CGキャラクターと人とのインタラクションにおける行為のモデル化を行ってきた。これらの研究を通じて、人間どうしの円滑なインタラクションに見られる規則性が、必ずしも人対エージェントの円滑なインタラクションに見られるものと同じであるわけではないことが明らかになってきた。

人間どうしのインタラクションをエージェント相手のインタラクションへ応用するとき、確かめなければならない課題として、「果たしてエージェントが身体を持ったとして、人間はその物に対して社会的インタラクションを行うのであろうか」という問題がある。現状の技術では人間同様の振る舞いをできるエージェントは存在せず、どのエージェントも人間と同等の発話や動作の認識・生成ができるわけではない。我々が作成している対話型ヘルプエージェントシステムも同様であり、そういったエージェントに対して、人は人に対するのと同じ社会的規範に則った振る舞いをするかどうかについては、大いに疑問の残るところである。

そこで、本研究では人対人、人対身体をもったエージェントの非言語行為を含む行為のやりとりを観察し、それぞれのインタラクションにおいて用いられている社会的規範を明らかにする。第2章では、人と人工物のインタラクションに関わる先行研究を整理し本研究の位置づけを示す。第3章では人どうしのインタ

† Verbal/Nonverbal Behavior Allocation Rules in Social Interactions between Users and Conversational Agents
Mika ENOMOTO and Yukiko NAKANO

*1 東京工科大学
Tokyo University of Technology

*2 成蹊大学
SEIKEI University

ラクションに関わる先行研究のうち、行為の配置規則の雛形として、言語のやりとりのための社会的規範として提唱されている話者交替システムについて概観する。そして、人対人、人対エージェントの対話資料(第4章)を用いて、まずそれぞれの対話の基礎的特徴を明確にする(第5章)。そして、第6章では、人間どうしのインタラクションから行為の配置規則を定式化する。第7章では、人対エージェントのインタラクションにおいて、この規則のどの部分が遵守され、どの部分が遵守されないかを明らかにする。そして、遵守されない規則の部分が、人が人とは違うエージェントとの対話への対処として選択した、円滑なインタラクションのための方便であることを明らかにする。

2. 人と人工物の社会的インタラクションに関わる先行研究

人対エージェントのインタラクションは、人対機械(人工物)のインタラクションであり、必ずしも人対人のインタラクションで用いられる社会的規範が守られるわけではないことは明らかである。人対人と人対エージェントのインタラクションの異同に関わる先行研究はこれまで盛んになされてきている。それらの研究をまとめると、大きく二つに分けられる。人工物に対して対人的認知がなされるかという社会心理学的な認知研究と、人工物に対してどのような行動が取られるかという行動科学的研究である。そして、両研究の下位区分として、研究対象が(1)人とは全く違う見かけをもつ人工物、(2)人と類似した見かけをもつ人工物、(3)人とそっくりな見かけをもつ人工物という3つに分かれる。ここで言う人間との類似度とはその動きやアピランスがどの程度人間に類似しているかを指す。コンピュータのハードウェアなどは、明らかに(1)の人とは全く違う人工物といえる。また、コンピュータ上のCGキャラクターなどは人と同じようにな動きができるよう口や目や耳、手足といった表現デバイスを持たせることが可能であるが、その身体は2次元上に存在する厚みの無い存在である。従って、(2)の人と類似した人工物である。(3)の人間と酷似した人工物とは、RepleeQ2(山森ほか、2007)など、人間に酷似した皮膚を持ち、人の身体動作と同様の動作ができるアンドロイドを指す。ただし、認知研究に関してはこの3つの下位部門それぞれに分類される研究が行われているが、行動科学研究では、人と酷似した人工物とのインタラクションは扱われていない。それぞれの部門に代表的な研究を以下に挙げる。

■認知研究

人とは異なる人工物 Nass et al.(1994)は、人とは明

らかに異なる人工物(コンピュータのハードウェア)に対しても、個性性の帰属がなされることを証明し、物理的実態としてのハードウェア一つがインタラクションの対象とみなされることを明らかにしている。竹内らも、同一のハードウェアなら、タスクを共に行ったソフトウェアとは異なるソフトウェアに対しても批判的な態度が控えられることを示しており、Nassらの仮説は支持されている(竹内・片桐, 1999)。

人と類似した人工物 竹内・片桐(ibid.)の研究と同じ実験をしたとき、ソフトウェアが人と類似したエージェントの場合、ハードウェア自体ではなく個々のエージェントを対象として、インタラクションの相手としての対人的帰属がなされることが明らかにされている(竹内・片桐, 2000a)。また、たとえ同一のハードウェア内に構築されているエージェントでも、ユーザに同意を示すエージェントに対してのみユーザの好感が与えられることから、人と類似したエージェントに関しては固体識別がなされることも示されている(竹内・片桐, 2000b)。

人と酷似した人工物 人と酷似した人工物を追求すると、「不気味の谷現象」(Mori, 1970)と呼ばれる現象が生じることが知られている。ロボットのアピランスがある程度人間らしくなるにつれ、より好感的、共感的存在として評価されるが、あまりに人間に似せすぎた人工物になると返って人に奇妙なものとして扱われるとされる。この理由としては、(1)対象が人間らしくなると返って非人間の特徴が目立つ、(2)不自然な振る舞いが精神的・身体的病気を思い起こさせる、などがある。ただし、本当に人間に等しい人工物ができたときに、不気味さが谷から回復するのか、という反論もなされている。

■行動科学的研究

人とは異なる人工物 寺田ほか(2007)は、動く椅子に椅子自身の意図性を帰属させた場合、動作の頻度が多くなったり、動作時間が長くなることを報告している。これは人とは全く異なる人工物に対しても、人間が何らかのインタラクションを試みようとすることを明らかにしている。

人と類似し人工物 Cassel & Thorisson(1999)は、人対エージェントのインタラクションにおいて、うなずきなどの非言語情報がエージェントによって使用されたときに円滑なインタラクションが生じることを示している。また、Takeuchi et al.(2001)は、エージェント間のインタラクション

において権威を付与されたエージェントに対して、人もそのエージェントの発言に従うようになることを実証している。また、Nakano et al. (2003)は、人対人と人対エージェントの発話中における相手への視線の向きに着目し、人工物であるエージェントに対して人も同様の発話地点で視線が向けられることを示している。これらの研究は、人対人のインタラクションにおいて用いられる方略が、人工物相手にも用いられることを示唆する。一方、林ほか(2007)は、人対人、人対機械のチャットによる相互作用を比べ、相手が人であると教示した場合と機械であると教示した場合の印象評定と対話全体の文字数をカウントしているが、課題が複雑になると教示の如何に関わらず、人対人の方が相手との協調度ややりとりの文字数が増えることを報告している。これは、人対人のインタラクションにおいて用いられる方略が、人工物相手にも用いられることに対する反証となっている。

以上が人と人工物とのインタラクションに関する研究の概観である。本研究は、この中の行動科学的研究であり人と類似したエージェントを扱う、人と類似したエージェントとのインタラクションと本当の人のインタラクションの異同に焦点を当てる。ただし、その差異を見出すことによって何を旨とするかという工学的パラダイムにおいては、従来研究と異なる立場を取る。従来は、エージェントに人間の行動モデルを実装し、より円滑で自然なインタラクションが可能のように、エージェントをより人間らしく改善していくことを目的としていた。そういった技術開発の根底には、やはり人間らしいエージェントの方が円滑なインタラクションを引き起こせるだろうという想定があったように思われる。人の行動モデルを構築するにせよ、人とエージェントとのインタラクションを比較するにせよ、最終的な目標は人らしいエージェントの設計を目指すことになる。従って、対人とは違う反応を人に引き起こしてしまうのは、エージェントの実装上の不備であり改造が必要な出来事として取り扱われてきた。

しかし、現状において実装できる人間らしいエージェントには、円滑で自然な発話や動作が困難である。にもかかわらず、人はそういったエージェントと対話する機会を得れば、何らかの形でインタラクションを行う。人は見慣れぬエージェントに対してどのような方法を用いてインタラクションを行うのであろうか。これまでの研究では、インタラクションの細部を観察し、あまり自然とは言えない対話相手に対して、

人がどのような方略を用いて対話を行っているかの研究が手薄であった。総発話数や対人印象を測定するだけでは、エージェントを相手にしたときの人間のインタラクション・ルールを明らかにすることはできない。このルールが明らかにしなければ、人間どうしのインタラクションから得られた知見を生かした対話エージェントの実現は不可能である。現状低コストで実現できるレベルのエージェントを相手にしたとき、人はどのような方略を選択するのであろうか。

本研究では、現状において実装できる人間らしいエージェントには、円滑で自然な発話や動作が困難であるという前提に立つ。そして、そういったエージェントと共に対話という活動を成し遂げるために人間が用いている方法を明らかにする。この方法を明らかにすることは、エージェントに人が自身に対してはどのような方法に基づいてインタラクションをしようとするのかの知識を授けることを可能にする。この知識は、人対人のインタラクションに比べれば、一見非円滑にみえるインタラクションも、エージェントが相手の場合には最も円滑なインタラクションが行われていることをエージェントに判断せしめることになる。人間どうしの対話であればトラブルと受け取られる出来事も、エージェント相手の対話では人にはそれがトラブル視されない些細な出来事であるという認識をエージェントにもたせることができる。本研究の工学的パラダイムは、対人とは異なる方法によって達成されるインタラクションの方法を実装したエージェントの設計を目指すという所にある。

3. 人どうしの社会的インタラクションに関わる先行研究

人間どうしの社会的インタラクションに関わる研究は、社会学の一派として誕生した会話分析に一日の長がある。会話分析は、H. ガーフィンケルやE. ゴフマンらによって創設されたエスノメソドロジーより出でている。エスノメソドロジーは、成員達が具体的な状況において行う推論の方法、解釈の方法、コミュニケーションの方法、相互行為の方法、会話をするための方法を記述・分析する学問体系である。H. サックスやE. シェグロフはこれを会話分析という体系へ展開し、会話における相互行為そのものを研究することで、そこに見られる規則性を発見し記述することを目指した(Turner(1974), サーサスほか(1989)などを参照のこと)。ここでいう規則性は例えて言うならゲームのルールである。重要なのはこのルールが、社会の成員の活動の中で、活動を通じて達成される様々な方法であり、それを生み出した成員にとって意味がある

ものである、という点である。このルールは研究者が先天的に考案した分析概念ではなく、社会の成員が状況に応じて用いているものである。3.1で詳述する話者交替システムがこのルールの代表的なものである。成員たちは通常このシステムに従うが、時にこれを破る。しかし、このシステムが破られた時、発話構成の中にピッチの上昇や音の引き伸ばしが入り込む。これらのイレギュラーな発話のデザインがなされるということそれ自体が、話者交替システムに則ってインタラクションを行うことを社会の成員たちが達成しようとしていることの証拠になる。

本研究で最終的に記述しようとしている行為の配置規則はこういった会話分析の概念を取り入れたものである。我々が提唱するルールは、機械的アルゴリズムとしてのルールではなく、対話の参与者達が言ったり、行ったりすることの中で運用している会話のルールである。比喩的に言えば、将棋のルールを知らない人間が、そのプレーを観戦することで、どのようなルールに則ってそのプレーが行われているのかを発見しようというものである。

では以下に会話分析家が発見した言語チャネルにおける行為のやりとりに関わる社会的規則である話者交替システムを紹介する。

3.1 話者交替システム

Sacks et al. (1974)は、会話を観察し、非常にシンプルな現象を発見した。それは、**話している人は常に1人**という現象である。何人の会話参加者がいようが、話し手は任意の1人に決められている。そしてこれを達成するためには、誰がどの順でどの程度の長さ話すかを前もって決めておかなくても、順番に1人ずつ話すことを可能にするための社会的規範が存在するのであろうと彼らは考えた。この規範が、話者交替システムである。話者交替システムは、ターン構成要素とターン配分要素という2つの成分とその使用ルールから成る (Sacks et al., *ibid.*)。

ターン構成要素 ターン構成単位 (turn constructional unit: TCU) から構成される。TCUは、文、節、句、単語など様々なタイプの単位から成り、このタイプはTCUがどこで完結するかについての投射をもたらす。話し手は話し出すとともにこの単位を用い、少なくとも1つの単位が完結するまで話す権限を持つ。この単位の終わりには話者移行適格場 (transition relevant place: TRP) がある。TRPとは、相手の発話を邪魔せずに次話者が発話を開始できる場所である (詳細は榎本 (2008))。

ターン配分要素 ターン配分要素は、(a) 現話者によ

る次話者選択テクニックと (b) 次話者による自己選択テクニックに分かれる。(a) には、現話者が呼びかけや視線を向けることによってある1人の聞き手を発話の宛先として選んでいる状態で、隣接ペア第一部分 (質問や要求などの言語行為) を用いるという方法がある。(b) には、どの聞き手より先に話し出すという方法がある。

円滑な話者交替はこのシステムに則って、TCU末にあるTRPにおいて行われるとされている。そして、文、節、句、単語などの統語的な情報がこのTRPを聞き手に予測させると言われている。聞き手は発話内容がどんなに早く理解できたとしても、TCUが終了しうる所まで次のターンの開始を見合わせなければならない。これが社会的ルールである。そして、TRPにおいて聞き手が発話を開始したとき、元の話者は次発話が開始される直前に発話を終了しているので、常に話者が1人という状態が達成できるとされる。

ただし、聞き手の次の行為が言語チャネルではなく非言語チャネルであった場合はどうなるのか。Sacksらの規則には、聞き手が非言語行動を行う場合の記載はない。同様に、話し手のターンが指差しなどの非言語チャネルから構成されていた場合の記載もない。言語チャネルから言語チャネルへの交替が生じるとき、もし重複が生じたなら聞き取りに認知的障害が生じるため、1人一時を達成するための規範が存在することは肯ける。しかし、言語チャネルから非言語チャネル、非言語チャネルから非言語チャネルなど、マルチモーダルなチャネルに渡って行為者が替わるとき、必ずしも聞き手は先の行為が終わるのを待たなくても認知的競合は生じない。従って、行為者は常に1人という状態は達成されなくてもよさそうである。

3.2 行為の交替に関わるマルチモーダル研究

聞き手は非言語行動を開始するにあたって話し手の発話が終了するのを待たないこと、動作開始のための十分な情報が揃ったならすぐに行動を開始することが指摘されている (Spivey-Knowlton et al., 1998; Spivey-Knowlton et al., 1999)。Clark & Krych (2004) は、教示者の指示に従って操作者が対象物を操作するという課題を詳細に分析し、

- ・ 指示発話の途中で、操作者が対象物とその操作方法を理解するや否や手を動かし始め、
- ・ この動作を見て、教示者が発話を中断し、その動作に対する承認を与えたり、教示内容の修正をする

ことを報告している。このことは、聞き手が教示者の発話を理解するや否やTCUの途中であっても行動を

開始することを示している。この結果は、TCUが終了するまで次の発話の開始を見合わせるという規範が存在するとしたSacksらの話者交替システムが、非言語チャンネルを含むインタラクションにはあてはまらないことを示唆している。

しかし、榎本ほか(2007)は、教示者の指示に従ってユーザがパソコン操作を行う課題を分析し、

- ・ ユーザのマウスクリックという非言語行動は教示者の発話の切れ目で生じる

という結果を得ている。これは、聞き手が理解と同時に行動を開始し、聞き手の非言語行動と話し手の発話のパラレルに発生するというClarkらの先行研究の結果と相反するものである。この結果は、行為の種類によっては非言語行動もTRPで開始されていることを示唆するもので、パソコンのマウスクリックという行為に関していえば、話し手の発話と聞き手の非言語行動が交互に配置されることを意味する。Schegloff(2007)は、非言語行為を要求する発話に対してはこれに対する承認と要求された非言語行為を行うという2つの行為が対になることを指摘しているが、要求された非言語行為の内容によっては、その行為を完成させるために長い時間のかかるものもある。Clark & Klych (ibid.)が行ったような対象物を操作する(指定されたブロックを取ってきて組み立てる)といった非言語行為を行う場合には、ブロックを取ってくるまでに時間がかかる。彼の報告から実際に組み立てる行為が開始された時間はわからないが、「置いてください」という発話に対して、まさに「置く」という非言語行為を実行するまでには長い時間がかかる。これらの行為対を時間上に配置するためには、長い時間を要する非言語行為の方を早目に始めるしかない。一方、マウスクリックのように一瞬で行為を完成させられるようなタイプの非言語行為が要求された場合、その行為の実行を一瞬でできるので予め準備しておく必要が無い。このように非言語行為を実行するに要する時間の違いによって、いつその行為が開始されるかには違いがあるだろう。本研究の目的はパソコン操作課題における非言語行為を実装することにあるので、マウスクリックなどの一瞬で実行可能な行為がどのように配置されるかを扱うことにする。

ここでは、人と人、人とエージェントの対話の基礎的特徴を押さえた上で、パソコン操作に関わる人同士の非言語行為を含む行為の配置規則を明らかにすることから始めることにする。まず、人どうしのインタラクション研究から得られた話者交替に関する知見を下敷きにする。そして、従来の話者交替システムを拡張し、パソコン操作時にマルチモダリティを用いて行

われる人の社会的インタラクションを捉える社会的規範を記述する。そして、この知見をベースとし、人とエージェントのインタラクションがどの程度この規範に沿って行われるものであるかを分析する。

4. 対話資料

テレビパソコンの操作を課題とし、対話者の一方を教示者、もう一方を学習者とする対話を収録した。

対象として、ビデオを介した人間同士のコミュニケーションと、モーションキャプチャーを用いてアニメーションエージェントの動きを制御した擬似的なエージェント対ユーザのコミュニケーションを収集する。

4.1 条件

本稿で分析に用いる実験条件は以下の2種類である。

- ・ **Human-to-human (H-H)条件**：教示者が人間であり、学習者はパソコンのスクリーンとその前に立った教示者を対話相手として課題を行う。教示者は120インチの大型スクリーンに投影されたパソコン画面の前に立ち、これをキャプチャーした映像が別室に居る学習者が操作するパソコンのモニターに映し出される(図1左側)。学習者は、この映像を見ながら、映像内のスクリーンに映ったパソコン画面の操作を行う。なお、教示者側のスクリーンに投影されているパソコン画面は、ユーザが操作するパソコンからの出力を用いた。また、教示者側には、ユーザの上半身がキャプチャーされた画像も提示された。なお、音声はヘッドセットマイクを通じて相互に伝達された。
- ・ **Human-to-CG animation (H-C)条件**：教示者がCGアニメーションによる手足と顔を持った兎のキャラクターであり、学習者はこのCGキャラクターを対話相手として課題を行う。このCGキャラクターはWizard-of-Oz法を用いて操作されている。HH条件と同じ状態で教示を行う人間の動きがモーションキャプチャーによって取得さ

図1 H-H条件(左), H-C条件(右)の学習者側モニター

れ、その信号がアニメーション生成システムに送られ、CGキャラクターの動きが生成される。学習者は、別室にてこのCGキャラクターのアニメーションが合成されたパソコン画面を見ながらパソコン操作を行う(図1右側)。この条件においても、教示者側には、ユーザの上半身がキャプチャーされた画像も提示された。音声はヘッドセットマイクを通じて相互に伝達された。ただし、教示者の音声は音声変換ソフトウェア(Herium)により機械的な音声に変換されたものを用いた。

4.2 課題と実験デザイン

教示者役、学習者役の被験者は上記2条件の環境(H-C/H-H)において、番組予約、DVD作成という2つの実験課題のいずれかを行う。学習者役(20名)は両条件を1回ずつ行い(計2対話)、教示者役(10名)は両条件を2人の異なるユーザ役相手に2回ずつ(計4対話)行う。実験課題は条件間でカウンターバランスされており、収集した対話の合計は40対話である。本研究では、このうち5組20対話分を分析資料として用いる。

4.3 実施方法

実験の実施は以下の手続きで行われた。

1. 教示者役の被験者は、あらかじめ実験に利用するソフトのマニュアルが配布され、事前学習をするとともに、実験当日には、実験実施者により知識の再確認が行われた。
2. 学習者役被験者は実験実施の直前に来訪し、H-H/H-C条件の両条件において、番組予約と予約された番組のDVDへの書き込みを行うことが課される。学習者への教示は「課題をできるだけ自力で遂行し、わからなくなったら教示者に尋ねてください」である。
3. HC条件においては、教示者にCGキャラクターになりきって学習者と対話を行うよう指示した。

対話は、教示者役の全体像、及びズーム画面、学習者役の上半身像とズーム画面、被操作システムのモニタキャプチャー等合わせて7台のビデオカメラを用いて録画される。また、教示者役・学習者役の被験者の音声はヘッドセットマイクを用いて収録用パソコンに録音されている。

4.4 転記

1. 各参加者の音声ファイルから、wavesurfer 1.8.5を用い、パワー・スペクトログラム情報をもと

に、200ミリ秒以上の無音区間で区切られた有声区間(IPU¹)の切り出しを行う。

2. MultiTrans 1.1を用い、IPUを単位とする発話内容の書き起こしを行う。

4.5 言語行動のラベリング

4.5.1 ターン構成単位の付与

転記に対してターン構成単位(TCU)の境界のラベリングを行う。TCU境界は、概ね文の統語的境界と同一になる(榎本ほか, 2004)。そこで、日本語における統語境界として提唱されている節単位(高梨ほか, 2004)を参考に、表1に挙げた箇所をTCU境界として認定する。

4.6 非言語行動のラベリング

4.6.1 ジェスチャー

ジェスチャーの形態とフェーズについてはMcNeil (1992)の分類に基づき、次のようなラベリングを行う。

■ジェスチャーの形態

ポインティング モニターのある場所を指し示す動作。円を描くようにして、複数の対象を囲うような動作もポインティングとみなす。

トレース 画面中の文字をなぞるようにして指し示すジェスチャー

その他 上記以外のジェスチャー

■ジェスチャーのフェーズ

preparation : ジェスチャー開始からstrokeまでのジェスチャー準備区間。

表1 ターン構成単位の境界

絶対境界	用言・助動詞終止・命令形、終助詞が出現した箇所、および、これらに引用のト類が後続して言い終わっている箇所
強境界	ガ・シ・ケレド系並列節が出現した箇所
弱境界	トイウ・トカ、シ系引用節、タラ・ト・ナラ・レバ系条件節、ダノ・デ・ナリ系理由節、その他タリ・テカラ・テハ・テモ・テ、ヨウニが出現した箇所
格助詞止め	格助詞の後200ミリ秒以上のポーズが空き、次のTCUが開始されている箇所
体言止め	名詞の後200ミリ秒以上のポーズが空き、次のTCUが開始されている箇所
言いさし	フィラーや言い淀みなどでTCUが中断された箇所
言い止め	自発的に、あるいは他者の発話開始により発話がTCUされた箇所

1 この有声区間単位を間休止単位(Inter Pausal Unit; 以下, IPU)と呼ぶ。

stroke：ジェスチャーが完成した瞬間。

hold：stroke後のジェスチャー保持区間。

retract：stroke,holdからのジェスチャー撤回区間。

partial-retract：次のジェスチャーへ移行するための、stroke,holdからのジェスチャー撤回区間

4.6.2 クリック箇所

学習者のマウス操作ログデータより、ログ自動収集プログラムを用い、ユーザ役による以下の3種類のマウス操作のログを収集した。

ポインターの対象物停留時間 エージェントによって指差されている対象物へポインターが入ってから出るまでの時間帯

マウスクリック マウスクリックがなされた時間帯

5. 分析1：人一人、人対エージェントの対話の基礎的特徴

まず、人一人、人対エージェントの対話の傾向を調べるため、H-H条件、H-C条件における発話量、発話構成を調べる。

5.1 結果

人一人、人対エージェントの対話の発話量をまとめたものが表2である。

IPU継続時間について、H-H/H-Cという対話条件と教示者/学習者という対話役割を2要因を説明変数とする一般線形モデルへの当てはめを行った。分析対象としたのは5組20対話に見られた全発話であり、各話者による発話特性を誤差項としてモデルに含めるため混合効果モデルを用いた。個人特性とは各発話者独自の話し方の特徴を指し、各話者に属する発話群ごとに説明変数が従属変数にどの程度影響するかを考慮した。

モデルとして選択したのは以下の3通りである。

対話役割モデル $IPU \text{ 継続時間} = a \text{ 対話役割} + (\text{個人特性による誤差} + \text{その他の誤差})$

表2 H-H条件、H-C条件における発話量の違い

		H-H条件	H-C条件
IPU平均時間	教示者	1.193秒	1.147秒
	学習者	0.690秒	0.653秒
TCU平均時間	教示者	2.202秒	2.660秒
	学習者	0.944秒	0.896秒
IPU数	教示者	1585	992
	学習者	1419	827
TCU数	教示者	907	553
	学習者	1355	702

対話条件モデル $IPU \text{ 継続時間} = a \text{ 対話条件} + (\text{個人差特性による誤差} + \text{その他の誤差})$

対話条件・対話役割モデル $IPU \text{ 継続時間} = a \text{ 対話条件} + b \text{ 対話役割} + (\text{個人特性による誤差} + \text{その他の誤差})$

a, bは係数、()内は誤差項を表す。IPU平均時間は実際の継続時間を対数変換した値を用いた。

それぞれのモデルにおいて、各変数の効果があるか否か(a, bの係数が0と比べて有意に大きくなるか否か)を分析した。その結果、対話役割モデルでは、対話役割による効果があったが($t(4819) = 5.92, p < .001$)。対話条件モデルでは、対話条件による効果はなかった。両要因を含む対話条件・対話役割モデルでも、対話役割による効果のみ見られた($t(4816) = 5.92 < .001$)。また、変数の追加によってモデルの選択が向上するか否かを調べたところ、対話役割モデルと対話条件・対話役割モデルの間に差は無いが、対話条件モデルと対話条件・対話役割モデルの間には0.1%水準で差が見られた($\chi^2 = 22.35$)。すなわち、対話条件間には差は無く、対話役割差による差のみ存在するという結果である。

TCU継続時間についても同様に変数の効果があるか否か、変数の増加による説明率の向上が見られるか否かの検定を行った。結果は、IPU継続時間と同様のものではなかった。対話役割モデルにおける対話役割の効果は0.01%水準で有意であり($t(3514) = 6.665$)、対話条件モデルにおける対話条件の効果はない。対話役割モデルと対話条件・対話役割モデルの間に差は無いが、対話条件モデルと対話条件・対話役割モデルの間には0.1%水準で差が見られた($\chi^2 = 26.234$)。対話条件間には差は無く、対話役割差による差のみ存在するという結果である。これらの結果は、個々のIPUやTCUの長さに関しては、H-H条件とH-C条件の間に差がないことを表している。

IPU数に関して同様の分析を行った。ただし、IPU数値には実際の生起度数を用い、これがピアソン分布に従うと仮定した。今度は、対話役割モデルにおける対話役割の効果は無く、対話条件モデル、対話役割・対話条件モデルにおける対話条件の効果のみ有意であった($t(42) = 13.42, p < .001$)。モデル比較を行ったところ、対話役割モデルと対話役割・対話条件モデルの間にのみ差が見られた($\chi^2 = 183.96$)。すなわち、IPU数に関しては、対話役割による差は無く、対話条件による差のみ散在するという結果である。

TCU数に関しても同様の分析を行い、同じ結果が得られた。対話役割モデルにおける対話役割の効果は無く、対話条件モデル、対話役割・対話条件モデルに

における対話条件の効果のみ有意であった($t(42) = 13.63, p < .001$)。モデル比較を行ったところ、対話役割モデルと対話役割・対話条件モデルの間にのみ差が見られた($\chi^2 = 191.43$)。

次に、発話構成を調べるため、H-H条件/H-C条件と教示者/学習者の2要因別にターン構成単位(TCU)の境界として認定された基準がいずれの境界であったかを計上した(表3)。

H-H条件/H-C条件、教示者/学習者、TCU境界の3要因を説明変数とする混交効果モデルへのあてはめを行ったところ、各要因の効果が見られた。対話役割で $t(228) = -9.732 (p < .001)$ 、対話条件で $t(228) = 0.674 (p < .001)$ 、TCU境界で $t(228) = -15.562 (p < .001)$ という差が見られた。対話役割(学習者/教示者)別に χ^2 検定を行ったところ、境界種別の偏りは、教示者のTCU境界のみ有意であった($\chi^2(6) = 13.34, p < .05$)。表3中()内の+-という記号は残差分析によって偏りの多少が検出された箇所である。これを見ると、H-H条件では、弱境界、格助詞止め、言いさし、言い止めといった境界が多く用いられることがわかる。反対に、H-C条件では、絶対境界、強境界、体言止めが多く用いられる。これは、H-C条件では、文末形式が整った文や名詞の一語文が用いられていることを示す。一方、H-C条件では、言いさしや言い止めなどの文法形式が整わない人間らしい発話デザインになっている。また、節末も弱境界でTCUが終了することが多く、これから行う行為の理由や条件が説明される傾向にある。

以上の結果をまとめると以下ようになる。

- ・ H-H条件、H-C条件において、教示者/学習者の

表3 H-H条件、H-C条件におけるTCU境界の違い

		H-H条件	H-C条件
絶対境界	教示者	667(-)	411(+)
	学習者	1044	545
強境界	教示者	18(-)	13(+)
	学習者	59	14
弱境界	教示者	105(+)	49(-)
	学習者	21	13
格助詞止め	教示者	19(+)	6
	学習者	33	14
体言止め	教示者	33(-)	20(+)
	学習者	124	67
言いさし	教示者	25(+)	15(-)
	学習者	59	28
言い止め	教示者	40(+)	20(-)
	学習者	51	21

発話長に差はない

- ・ H-H条件では、学習者の発話が多い
- ・ H-H条件では、教示者のTCUが弱境界、格助詞止め、言いさし、言い止めといった境界によって区切られる

5.2 考察

図2は、H-H条件における対話例である。教示者(B)の発話内容を見ると、弱境界、強境界、格助詞止めなどが用いられ、それぞれのTCU末で学習者(A)の相づち、応答が挟まっている。こういった相づちや応答が学習者の発話量を増やしている。

図3はH-C条件における対話例である。教示者(B)の発話内容を見ると、強境界、強境界、絶対境界が用いられている。このようにH-C条件では、H-H条件と同じ教示者が教示を行っているにもかかわらず、行為の理由や条件といった説明がなされることは少なく、何を行うのかという行為の指示だけから発話が構成される傾向にある。そして、これらの各TCU末において学習者の応答がなされることなく、「クリックしてください」に対して、マウスクリックがなされるという反応のみが生じている。これは極端な例であるが、H-C条件においては学習者の相づちや応答がなされることは全般的に少なく、これが学習者の発話量を減らしている。

条件間における教示者の発話量や発話構成の違いは、実験を行うにあたって与えた指示(「HC条件においては、CGキャラクターになりきって学習者と対話を行ってください」)による影響が強くてたものと考えられる。実験後の感想で、教示者役被験者は、「CGキャラクターになりきるために、できるだけ言い誤りや言いなおしをしないように勤めたが、何箇所かはそういった表現になってしまい申し訳ない」と言う人が多かった。言い誤りや言いなおしをしないようにするため、H-C条件においては、発話量を減らし、行動の指示のみを行うという方略が取られたため、こういった条件間の差が生じたといえる。

条件間における学習者の発話量の違いは、エージェントと話すことへの拒否感が表れているのだと考えられる。実験開始前に、「本当にこの兎と話さなければなりませんか?」と問うた被験者も居た。ただし、実験にはWizard-of-Oz法が用いられたにも関わらず、対話相手が実は人間であることに学習者役被験者は全員気づいていた。気づいた理由を問うた所、「発話が流暢すぎる」「説明が適切すぎる」などの答えが返ってきた。そして、対話相手が本当は人間であるということはずぐにわかったそうである。にもかかわらず、H-

図2 事例1. H-H条件における対話. Aは学習者, Bは教示者を表す. 数値は対話内での経過時間を示す.

図3 事例2. H-C条件における対話. Aは学習者, Bは教示者を表す. 数値は対話内での経過時間を示す.

H条件と比べると, H-C条件では明らかに相手と対話することが避けられていた. これは, エージェントである画面中のCG-キャラクターというメディアを介した対話が与えた影響であると考えられる. そして, キャラクターとの対話を避けるために, 課題の実行に関わる最小限の質問に留めるという方法が用いられていた.

相手が人間であれば, 相づちや応答の不在は社会的規範への違反として受け取られる. 相づちは相手の発話を聞いている, 理解している証拠として聞き手から提示されるもので, これがなければ話し手は発話の継続を中断したりして相手の反応を待つといった措置が取られる. また, 応答は指示に対して隣接第二部分を形成するものであり, 応答の不在は問い直しや指示のやり直しといった有標な行為を伴うことが指摘されている(Sacks, 1973[1987]). これは, 隣接第一部分が用いられたなら, 速やかに第二部分を行うこと, という社会規範が働いたためである. しかし, 本研究で分析対象としたH-C条件における対話では, 明らかに相づちや応答の不在があるにも関わらず, これに対処する措置はとられていなかった. エージェントとの対話であればこういった規範を遵守しなくてもよいという対話のルールがここでは作られていたと考えられる. そして教示者も, 応答がなくても, どんどん次の指示へ移っていく. 応答がないにもかかわらず, 教示者はなぜ学習者とその指示を聞き漏らしていないか, 理解に失敗していないかを掌握できるのであろうか. その答えは, 学習者のマウス操作(ポインターの移動やマウスクリック)にあると考えられる. 教示者は学習者の理解状態をそのマウス操作から推し量ることが可能だったのであろうと考えられる. 図3にも表れているように, 通常応答が入る位置にマウスクリックが生じて

いる. まさに, 応答の変わりにクリックがなされているといえる. これによって, 学習者が指示対象を発見し, それにたいする操作方法を理解したことを教示者は知ることができるであろう. では, こういったマウス操作が生じる適切なタイミングはいつか, それは人対人の対話と同じ位置で生じるのであろうか.

6. 人対人の対話における行為の配置規則の定式化

本章では, 言語チャネルから非言語チャネル, 非言語チャネルから非言語チャネルなど, マルチモーダルなチャネルに渡って行為者が替わるとき, それらの行為が時間軸上にどのように配置するのかに関して, 人間どうしのインタラクションで用いられているルールを発見する. 指差しジェスチャや発話とユーザのマウス操作が交互に行われるのか, パラレルに行われるのかを明らかにし, それらの行為を時間軸上に配置するためのルールを定式化する.

6.1 出発点としての最も単純な仮説

まず, 行為の配置規則の最も単純な法則として, 行為者の各行為が時間軸上へ順番に配置されると仮定する. エージェントの発話と指差しが終了した直後に, ユーザのマウス操作であるポインター移動とマウスクリックが生じるとすると, 図4のようになる. 図の上側はエージェントの発話と非言語行動, 下側はユーザの非言語行動を表す. 行為者の各行為が時間軸上へ順番に配置されるなら,

1. TCU終了直後に,
 - (a) ポインターが対象物上へ置かれる
 - (b) マウスクリックが開始される

図4 行為の配置規則に関する仮説

2. 対象物への指差し終了直後に、

- (a) ポインターが対象物上へ置かれる
- (b) マウスクリックが開始される

と仮定できる。話者交替規則から類推するなら、先行する行為と次の行為の間隙は隙間なく配置されるため²、作業仮説は以下ようになる。

対話中での各行為が生じた時間は、

1. (a) TCU終了時間=ポインターの対象物停留開始時間
 1. (b) TCU終了時間=マウスクリック開始時間
 2. (a) 指差し終了時間=ポインターの対象物停留開始時間
 2. (b) 指差し終了時間=マウスクリック開始時間
- 本章ではこの仮説を検証する。

6.2 各行動間の隔たり時間の算出

教示者のTCUと指差し、学習者のポインターの対象物停留時間とマウスクリックそれぞれの時間的隔たりを以下のように算出した。

- (1) **TRP-pointer** TCU終了点からポインターの対象物停留までの隔たりを、[ポインターの対象物停留開始時間-TCUの終了時間]として算出した(計46箇所)。
- (2) **TRP-click** TCU終了点からマウスクリックまでの隔たりを、[マウスクリックの開始時間-TCUの終了時間]として算出した(計42箇所)。
- (3) **pointing-pointer** 指差し終了点から対象物へのポインターの対象物停留までの隔たりを、[ポインターの対象物停留開始時間-指差しのretract終了時間]として算出した(計49箇所)。
- (4) **pointing-click** 指差し終了点からマウスクリックまでの隔たりを、[マウスクリックの開始時間-指差しのretract終了時間]として算出した(計44箇所)。

² 話者交替の場合、発話と発話の隔たりは約1mora程度(70～80ミリ秒)である(榎本, 2007)

(1)～(4)の値が0以外であれば仮説が支持され、行為者の各行為は時間軸上へ順番に配置されていることになる。そこで、隔たり時間が平均0ミリ秒となるか否かを調べるため1標本 t 検定を行う。ただし、対話参与者による発話回数の違いを考慮するため、一般線形モデルによる検定手法を用いる。

6.3 TCUからの隔たり

H-H条件におけるTCU終了点からのポインターの対象物停留開始とマウスクリックまでの隔たり時間((1)と(2))が平均0となるか否かを調べた結果が表4である。(1)TRP-pointerにおける全体の平均値は738ミリ秒であり、SDは3.058である。(2)TRP-clickの平均値は170秒であり、SDは2.041である。

t 検定の結果、ポインターの移動に関しては、10%水準ではあるが、TCU終了時間>ポインターの対象物停留開始時間となり($t(45)=-1.638$)、仮説を支持しない。一方、クリックに関しては、有意差がなくTCU終了時間=マウスクリック開始時間となり($t(41)=-0.546$)、仮説を支持する。すなわち、仮説1.(b) TCU終了時間=マウスクリック開始時間のみが検証されたことになる。

図5は、教示者の行為と学習者の行為の生起位置を平均値を元に模式化したものである。本節では、教示者のTCU終了地点と学習者の対象物へのポインター停留開始地点、マウスクリック地点に注目されたい。学習者の対象物への滞留時間が教示者のTCUにやや入り込んでいる一方、マウスクリックはTCU終了直後に生じている。教示者の発話とユーザのマウス

表4 H-H条件におけるTCUからの行為の隔たり時間

	t 値	自由度	SD	p 値	平均値(秒)
(1)TRP-pointer	-1.638	45	3.05	<.10	-0.738
(2)TRP-click	-0.546	41	2.041	n.s.	0.17

図5 H-H条件の各行為間の隔たり

リックは交互に生じている。すなわち、TCUとマウスクリックのみが時間的にみると、順番に配置されていることがわかる。

6.4 指差しからの隔たり

H-H条件における指差し終了時間からのポインターの対象物停留開始とマウスクリックまでの隔たり時間((3)と(4))が平均0となるか否かを調べた結果が表5である。(3)pointing-pointerの平均値は277ミリ秒であり、SDは3.521である。(4)pointing-clickの平均値は857ミリ秒であり、SDは2.039である。

t 検定の結果、(3)pointing-pointerでは有意差が見られなかったが($t(48)=0.549$)、(4)pointing-clickでは有意差が見られた($t(48)=2.78$, $p<.001$)。指差しと対象物への停留は時間軸上へ順番に配置されるが、指差しとマウスクリックでは順番に配置されたとは考えられないほど時間的隔たりが大きいという結果であった。

各行為間の隔たり時間長(平均値)を見ると、ポインターが対象物へ入れられるのは指差しが終わるとほぼ同時(227ミリ秒)であるが、マウスクリックはそれよりも600秒近く遅れて857ミリ秒後に起こる(表5参照)。榎本(2007)の実験によれば、人間が刺激の提示音に即反応してボタン押しをした場合、 221.12 ± 36.76 ($N=18$)ミリ秒程度でのボタン押しが可能である。認知処理の複雑さや行為完結部への予測の正確さによってこの時間は多少の変化があると考えられるが、指差しが終わって277ミリ秒後にポインター移動が完了するということは、学習者が指差しの終了を見て速やかにポインターを移動し終えたことを意味する。逆に、それからマウスがクリックされるまでさらに600ミリ秒程度の隔たりがあるということは、ポインター移動が完了すると同時にクリックがなされるわけではないことを示す。マウスクリックは、指差し終了後しばらく経ってから行われるものと考えられる。

図5の教示者の対象物への指差しと学習者の対象物へのポインター停留開始、マウスクリックに注目されたい。指差しは教示者の発話が終了するやや手前で生じている。学習者のポインターの対象物への停留はやや発話に重複する形で生じているので、指差しとの隔たりに差は見られないが、マウスクリックは発話終了

後に行われているため、指差しとの隔たりが大きくなる。この結果から、教示者の指差しと学習者のポインター移動は時間軸上へ交互に配置されるが、教示者の指差しと学習者のマウスクリックは時間軸上へ交互に配置されるものではないといえる。

6.5 人間どうしのインタラクションにおける行為の配置規則の抽出

分析の結果、

1. 対象物への指差し終了点で、ポインターが対象物上へ置かれ、
2. TCU終了点で、マウスクリックが開始される。

という仮説が支持された。すなわち、教示者の<TCU>と学習者の<ポインターの対象物停留開始>、教示者の<対象物への指差し>と学習者の<マウスクリック>は時間軸上へ直線的に配置されるといえる。

この結果をもとに、教示者の<対象物への指差し>、<指差し時に発されたTCU>と学習者の<対象物へのポインター停留>、<マウスクリック>という行為の連鎖規則を定式化したものが以下である。

Rule1 指差しが終了するところで、マウスを対象物上へ置く

Rule2 指差しに伴う発話が終了するところで、マウスクリックを行う

この規則はインタラクション上のルールであり、学習者と教示者の双方が従うべき規則である。学習者が対象物を発見できずにいるとき、教示者は指差しを保持し続けたりやり直すことが望まれる。図6は対象物の発見に手間取り何度も指差しがやり直される例である。最初の教示者発話「ビューのボタン押ししてもらっていいですか」が終了点をむかえているのに、学習者のポインターが対象物上へ移動しない。教示者は指差しを下ろすことなく、もう一度同じ箇所を指差し、追加説明を行う。249ミリ秒付近でポインターが対象物上へ移動するまでこの状態が続く。学習者のポインターが対象物である「ビューボタン」へ入った途端に、教示者は指差しを撤回し、2度目のTCU終了点「押してください」をつくる。ここでクリックが起こっている。

この例のように、ポインター移動が完了したのを見て指差しを撤回する方が、Rule1に従うためには確実である。ポインターが対象物に入る前に指差しを撤回してしまうのは、後にポインター移動が本当に伴うかどうかかわからず一種の賭けであるように思える。しかし実際には、教示者は、学習者がポインターを対象物へ移動させられるかどうかは早期の段階で予期できる

表5 H-H条件における指差しからの行為の隔たり時間

	t 値	自由度	SD	p 値	平均値(秒)
(3) pointing-pointer	0.549	48	3.521	n.s.	0.277
(4) pointing-click	2.78	43	2.039	<.10	0.857

図6 事例3. 対象物の発見に手間取るケース

と考えられる。Murata et al. (ibid.)は、学習者の操作するポインターの動きから、教示者は学習者の理解状態を推測し、追加説明や指差しの繰り返しを行うとしている。すなわち、実際にポインターが対象物内に入る随分前から、ポインターのその後の動きを予測することが可能なのである。従って、実際にポインターが対象物に入っていないくても、このRule1に違反することがないという確信の下、指差しを撤回できるのである。

7. 人対エージェントの対話において行為の配置規則は守られるか

ではいよいよ、人対エージェントのインタラクションを観察し、人と人のインタラクションにおいて利用されている行為の配置規則が、エージェントとのインタラクションにおいてどこまで遵守されるのかを見ていく。

7.1 各行動間の隔たり時間の算出

H-C条件で収録された対話から、教示者のTCUと指差し、学習者のポインターの対象物停留時間とマウスクリックそれぞれの時間的隔たりを6.2同様の方法で算出した。

- (1) **TRP-pointer** TCU終了点からポインターの対象物停留までの隔たり時間(計46箇所)。
- (2) **TRP-click** TCU終了点からマウスクリックまでの隔たり時間(45箇所)。
- (3) **pointing-pointer** 指差し終了点から対象物へのポインターの対象物停留までの隔たり時間(55箇所)。
- (4) **pointing-click** 指差し終了点からマウスクリックまでの隔たり時間(48箇所)。

(1)～(4)の値が0以上であれば仮説が支持され、行為者の各行為は時間軸上へ順番に配置されていることになる。

まず、隔たり時間が平均0ミリ秒となるか否かを調べるため1標本 t 検定を行う。もし(1)～(4)のいずれかで有意差が検出された場合には、各行動の隔たり時間を従属変数とし、H-H/H-C条件およびどの行為間の隔たりであるか(隔たり要因)を独立変数とする2

要因分散分析を行い、どの要因が影響しているのかを調べる。

7.2 TCUからの隔たり

H-C条件では、対象物の同定に手間取り非円滑なユーザの反応がなされたケースが数例あり、相当長い隔たり時間が観測される。ここでは、それらを分析から除外し、円滑な行為交替が行われた箇所のみを対象とする。以下の分析では、対象となる行為の隔たり時間の(平均値±SD)ミリ秒以内に収まるデータを用いる。(1)TRP-pointerの平均値は1.152秒であり、SDは51.072であった。そこで、平均値±SDに収まらない4ケースを除外したところ、表6に挙げるように、有効データは42ケースとなり、平均値1.141、SD=2.850となる。(2)TRP-clickは、全体の平均値が2.655秒、SD=53.146であった。平均値±SDに収まらない3ケースを除外したところ、有効データ数は42となり、平均値は0.012秒、SD=1.617となった。

これらの有効データに対して、1標本 t 検定を行ったところ、(1)TRP-pointerは0.1%水準で有意となり($t(41)=-3.219$)、TCU終了時間がポインターの対象物停留開始時間より長くなる。これは、仮説を支持しない結果である。一方、(2)TRP-clickに関しては有意差がなくTCU終了時間とマウスクリック開始時間が等しくなり、仮説を支持する。すなわち、仮説1.(b)TCU終了時間=マウスクリック開始時間のみが検証されたことになる。H-C条件でも、マウスクリックに関してのみ、TCU終了時間とマウスクリック開始時間が近似し、教示者の発話と学習者のマウスクリックは順番に配置されるという仮説が検証されたことになる。また、ポインターの対象物停留開始に関しても、H-H条件同様、有意差が検出されなかった。

H-H/H-C条件とTCU終了時間からのポインターの停留開始とマウスクリックまでの隔たり((1)と(2))

表6 H-C条件におけるTCUからの行為の隔たり時間

	t 値	自由度	SD	p 値	平均値(秒)
(1)TRP-pointer	-3.219	41	2.85	<.0.01	-1.416
(2)TRP-click	0.051	41	1.617	n.s.	0.012

を独立要因とする分散分析の結果，行為の隔たり要因の主効果のみが有意であった ($F(1, 168)=9.588, p<.005$)．要因ごとの平均値と標準誤差を図7に示す．チューキーのHSD検定を行ったところ，H-C条件の(3)TCU-pointer時間とH-H/H-C条件両方の(2)TRP-clickとの間に差が見られた ($p<.01$)． t 検定の結果同様，H-C条件のポインターの対象物停留開始がTCU内部に深く食い込んで発生しているために，このような結果となる．

図8は，H-C条件における教示者の行為と学習者の行為の生起位置を平均値を元に模式化したものである．教示者のTCU終了地点と学習者の対象物へのポインター停留開始地点，マウスクリック地点に注目されたい．教示者のTCUとマウスクリックは順番に配置され，H-H条件と同じ結果を示している．ところが，ポインターの対象物への停留開始時間は大きくTCU内部へ食い込んでいる．H-C条件では，教示者の発話と対象物への停留時間は順番に配置されるのではなく，平行に進行していた．

図7 TRPからの隔たり時間の平均値と標準誤差

図8 H-C条件の各行為間の隔たり

7.3 指差しからの隔たり

指差しからの隔たりに関しても，対象物の同定に手間取り非円滑なユーザの反応がなされたケースが数例あり，相当長い隔たり時間が観測される．

(3)pointer-pointerの平均値は-1.181秒であり，SDは47.466であった．そこで，平均値±SDに収まらない5ケースを除外したところ，有効データは50ケースとなり，表7に挙げるように，平均値-2.537秒，SD=3.538となる．(4)pointer-clickは，全体の平均値が1.562秒，SD=49.88であった．3ケースを除外したところ，有効データ数は44となり，平均値は-0.792秒，SD=2.872となった．

これらの有効データに対して，1標本 t 検定を行ったところ，表7のような結果が得られた．

H-C条件(3)pointing-pointerでは非常に大きな差が見られ(0.1%水準)， t 値も大きな負の値を示していた ($t(49)=-5.071$)．これは，指差し終了時間がポインター移動完了開始時間より大幅に遅いことを意味する．すなわち，指差しの終了を待たずに，ポインターが移動が完了されているということである．また，(4)pointing-clickでも10%水準ではあるが有意差が見られた ($t(43)=-1.829$)．これらの結果は，指差しに対してはポインターの対象物への停留もマウスクリックも順番に配置されていないことを示している．

図8の教示者の対象物への指差しと学習者の対象物へのポインター停留開始，マウスクリックに注目されたい．H-C条件では，指差しは教示者の発話の終了後まで持続される．このことが，H-H条件とは違い，教示者の指差しとマウスクリックの間の差を解消させている．ポインターの対象物への停留開始に関しては，教示者の指差し内部(発話末から平均2537ミリ秒付近)に深く入り込んでおり，教示者発話と対象物へのポインター移動は平行に進行している．

H-H/H-C条件と指差し終了時間からのポインターの対象物への停留開始とマウスクリックまでの隔たり((3)と(4))を独立要因とする分散分析の結果，各要因の主効果のみが見られた．H-H/H-C条件による要因の主効果は $F(1, 183)=5.899, p<.01$ であり，隔たり要因の主効果は $F(1, 183)=6.62, p<.001$ であった．要因ごとの平均値と標準誤差を図9に示す．チューキーのHSD検定を行ったところ，H-C条件の

表7 H-C条件における指差しからの行為の隔たり時間

	t 値	自由度	SD	p 値	平均値(秒)
(3) pointing-pointer	-5.071	49	3.538	<.001	-2.537
(4) pointing-click	-1.829	43	2.872	<.10	-0.792

図9 指差しからの隔たり時間の平均値と標準誤差

(3)pointing-pointer時間が -2537 ± 436 ミリ秒と他より抜きんでて値が小さく ($p < .01$)、これが結果に決定的な影響を与えていた。H-H条件の(3)pointing-pointer、H-H/H-C条件の(4)pointing-clickどうしの間には差が見られなかった。この結果は、指差しの終了時間の差を考慮すると、H-C条件における学習者のポインターの対象物への停留開始だけが明らかに早いタイミングでなされているといえる。

7.4 エージェントとのインタラクションに適用される行為の配置規則

本節の分析を通じ、CGキャラクターの教示者に対して学習者は、マウスをポインターを対象物へ置くと、相手を妨害するかどうかを気にせず、相手の占有すべき時間へ大幅に割り込んでいることが明らかになった。すなわち、人間どうしのインタラクションにおいては守られる

1. 対象物への指差し終了点で、ポインターが対象物上へ置かれるという規則が、CGキャラクター相手には守られないことがわかった。

一方、マウスクリックに関してはH-C条件においても、

2. TCU終了点で、マウスクリックが開始されるという規則が遵守されることが明らかになった。

なぜエージェントとのインタラクションでは、ポインター移動の開始が早められるのか。第5章で、人対エージェントの対話では、学習者の発話量が少ないという基礎的特徴があることを指摘した。そして、学習者の発話には相づちや応答が少ないのに、教示者はなぜ学習者が指示対象を発見し、それにたいする操作方法を理解したことを知ることができるであろうかという疑問を提示した。これに対する解答がここにある。

学習者は相づちを打つ代わりに、教示者が指差しした箇所を発見するや否やポインターをそこへ移動させる。人間どうしのインタラクションでは、ポインターの移動は指差しが終了するまで待たれるが、エージェントとのインタラクションでは、これに違反してでも先にポインターが移動される。この動作によって教示者は学習者が指示対象を発見したことを知り、スムーズに次の指示へと移れるのである。

図10は学習者と教示者の各行為の生じた時間帯を表したものである。教示者は画面中の[録画予約]というボタンを指差しながら、「その状態で」とTCUを開始する。一回目の指差しが完了した時点ではユーザのポインター移動が起こらない。そこで、再度指差しを行い、「録画予約を」とTCUを継続する。ここで、学習者が操作するポインターが動き始め399ミリ秒付近で、対象物の上へ置かれる。これを見た教示者は指差しを止め、「一回押してください」とTCUを完成させる。するとユーザのマウスクリックが起こる。この間、学習者は一切発話を行っていないが、その理解状態はマウスポインターの動きに反映されている。ポインターの動きが止まっているという情報を利用して、教示者は発話を中断したり、ジェスチャーをやり直したりと色々な工夫をしている。そして、ポインターが指示対象へ向かうという情報を利用して、TCUを完結させている。このように、学習者は、対話相手として好ましくないCGキャラクターと交わす言葉を少なくするかわりに、マウス操作に含まれる情報を逸早く教示者に伝えている。また、教示者もこの情報を利用して発話やジェスチャーを組み立てている。これが、エージェントを介した会話を行うために、参与者たちが生み出した社会的インタラクションのルールである。人対人のインタラクション・ルールと比べると、一見ルール違反のように見える行動も、精緻に組み立てられたインタラクションの一方略であるといえる。

一方、マウスクリックに関しては、人対エージェントの対話においても、人対人と同じルールが適用されている。違いは、人対人のときにはマウスクリックと同時に「はい」などの応答発話が生産されるが、人対エージェントの時にはそれが無いことが多いという点だけである。なぜ、クリックに関してはTCU内部に入り込むことなく、TCUが完結した地点に配置されるのだろうか。この理由の一つとして、マウスクリックという行為がこの会話運用にもたらす影響力の強さがあると考えられる。クリックは画面の遷移をもたらす。会話を次のステップへと導く。図10の例で言えば、クリックの後、録画予約画面へとモニターが変化する。もし、教示者のTCUが終わる前にこの遷移

図10 事例4. H-C条件における学習者(A)と教示者(B)の行為の配置. 上から順に学習者のマウスクリック, ポインターの滞留した対象物名, 教示者の発話, 指差しの各フェーズを表す.

が行われたとしたら, 教示者はTCUを中断し, 次の説明や指示を開始しなければならない. これは, 明らかに話者交替システムへの違反である. 従って, そのような深刻な事態を引き起こす力のある行為に関しては, たとえそれがCG キャラクターに向けられてなされた非言語行動であろうと, 相手の行為を遮らない位置に配置することが強く指向されているのだと考えられる.

8. まとめ

本研究では, パソコン操作課題において人とエージェントが対話するとき, どのように言語・非言語行為をインタラクションの中で配置するのかを明らかにした. まず, 人対人と人対エージェントの対話の基礎的特徴を観察し, 人対エージェントの対話では人の発話量が少なく, 相づちや応答が稀にしか差し挟まれないことを見た. この振る舞いは一見, 相手発話の理解や承認を伝達することを妨げ, インタラクションの円滑さを損なう行為であるように見える. しかし, 人対人の行為の配置規則において交互に配置される指差しとポインター移動が, 人対エージェントのインタラクションでは, パラレルに配置されることを見た. これは, 理解の証拠を言語行為ではなく非言語行為によって相手に伝達するもので, 円滑なインタラクションを補償するものであることを示した. 一方, マウスクリックに関しては人に対するのと同じように, 指示発話の後に配置されていた. これは, たとえエージェント相手であろうとも遵守される規則があることを示している.

このように, エージェントを相手にした時のインタラクション方略に人相手のときとの異同があるということは, 人が人工物を相手にしたときにインタラクションルールの改変が一部行われるということを示すものである. 本研究では, 研究対象をCGキャラクターであるエージェントに限定しているが, ユーザの課題遂行を援助する熟練者や専門家の代理人という意味では, 同様の役割をヒューマノイドロボットに担わせることは可能である. 将来CGキャラクターと同程

度の動きが可能なロボットが出来た暁には, 本研究の延長線上の議論が展開できるであろう. 人間の行動モデルをベースにヒューマノイドロボット/エージェントを実装するとき, ヒューマノイドロボット/エージェントに適したインタラクション・モデルを組み込むことが必要になると考えられる.

謝辞

本研究は, (株)日立製作所中央研究所と東京農工大学との共同研究として2006年~2007年に実施されたものである. 本研究の遂行にあたり, 有益な議論とご協力をいただいた(株)日立製作所中央研究所 佐川浩彦氏, 朝康博氏に深く感謝いたします.

参考文献

- Cassell, J., & Thorisson, K. (1999). The Power of a Nod and a Glance : Envelope vs. Emotional Feedback in Animated Conversational Agents. *Applied Artificial Intelligence*, 13, 519-538
- Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50, 62-81.
- 榎本美香. (2007). 日本語におけるターン構成単位の認知メカニズム. *社会言語科学*, 9 (2), 17-29.
- 榎本美香. (2008). 非言語チャネルによる社会的行為の隣接対に関する記述. *人工知能学会研究会資料, SIG-SLUD-A703*, 87-92.
- 榎本美香・赤木康宏・有本泰子・村田和義・朝康博・佐川浩彦・中野有紀子. (2006). 操作課題における情報授受のための行為者交替ルール-wizard-of-oz法を用いたヒューマン・エージェント会話実験を通じて. *HAIシンポジウム2006* (p. 2B-2). 東京
- 榎本美香・石崎雅人・小磯花絵・伝康晴・水上悦雄・矢野博之. (2004). 相互行為分析のための単位に関する検討. *人工知能学会研究会資料, SIG-SLUD-A402*, 45-50.
- 林勇吾・三輪和久・浦尾彰. (2007). 人間と人間, 人間と対話エージェントとの相互作用に関する実験的検討の一考察. *HAIシンポジウム2007* (p. 1E-1). 東京.
- McNeil, D. (1992). *Hand and mind, what gestures reveal about thought*. Chicago : The University of Chicago Press.
- Mori, M. (1970). The uncanny valley. *Energy*, 7 (4), 33-35.

- Murata, K., Enomoto, M., Arimoto, Y., & Nakano, Y. (2007). When Should Animated Agents Give Additional Instructions to Users? – Monitoring user's understanding in multimodal dialogues –. In *Proceedings of 2007 International Conference on Control, Automation and Systems (iccas 2007)* (pp. 733–736). Seoul, Korea.
- 中野有紀子・村田和義・榎本美香・有本泰子・朝康博・佐川浩彦. (2007). ペイジアンネットワークを用いたヒューマンエージェントインタラクションのモデル化. 第21回人工知能学会全国大会 (p. 2D5-9). 宮崎.
- Nakano, Y., Murata, K., Enomoto, M., Asa, Y. A. Y., & Sagawa, H. (2007). Predicting Evidence of Understanding by Monitoring User's Task Manipulation in Multimodal Conversations. In *Proceedings of the ACL 2007 Demo and Poster Sessions* (pp. 121–124). Prague, Czech Republic.
- Nakano, Y., Reinstein, G., Stocky, T., & Cassell, J. (2003). Towards a model of face-to-face grounding. In *Proceedings of the 41st annual meeting of the Association for Computational Linguistics* (pp. 553–561). Sapporo, Japan.
- Nass, C., Steuer, J., & Tauber, E. (1994). Computers are social actors. In *Proceedings of the CHI conference* (pp. 247–252).
- サーサス G.・ガーフィンケル H.・サックス H.・シェグロフ E. (1989). 日常性の解剖学. 東京: マルジュ社. (北沢裕・西坂仰 訳)
- Sacks, H. (1973 [1987]). On the preferences for agreement and contiguity in sequences in conversation. In G. Button & J. R. E. Lee (Eds.), *Talk and social organization* (p. 54–69). Multilingual Matters.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735.
- Schegloff, E. A. (2007). *Sequence organization in interaction*. Cambridge University Press.
- Spivey-Knowlton, M. J., Tanenhaus, M. K., & Eberhard, J. C., K. M. and Sedivy. (1999). Eye movements and spoken language comprehension; effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45(4), 447–481.
- Spivey-Knowlton, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (1998). Integration of visuospatial and linguistic information; language comprehension in real time and real space. In P. Olivier & K. P. Gapp (Eds.), *Representation and processing of spatial expressions* (pp. 201–214). Mahwah, NJ; Laurence Erlbaum Associates.
- 高梨克也・内元清貴・丸山岳彦. (2004). 『日本語話し言葉コーパス』における節単位認定. (『日本語話し言葉コーパス』DVD-ROM, 国立国語研究所・情報通信研究機構, <http://www.kokken.go.jp/katsudo/seika/corpus/public/manuals/discourse.pdf>)
- 竹内勇剛・片桐恭弘. (1999). 社会的な人-コンピュータインタラクションにおける個性性の帰属. 情報処理学会論文誌, 40(2), 623–631.
- 竹内勇剛・片桐恭弘. (2000a). ユーザの社会性に基づくエージェントに対する同調反応の誘発. 情報処理学会論文誌, 41(5), 1257–1266.
- 竹内勇剛・片桐恭弘. (2000b). 社会的インタラクションにおけるエージェントの個性性の帰属. ヒューマンインタフェース学会研究報告集, 2(3), 91–98.
- Takeuchi, Y., Katagiri, Y., & Takahashi, T. (2001). Learning enhancement in Web contents through inter-agent interaction. In *Proceedings of INTERACT 2001*.
- 寺田和憲・社本高志・梅海櫻・伊藤昭. (2007). 意図的人工物. 第21回人工知能学会全国大会 (p. 2D5-7). 宮崎.
- Turner, R. (1974). *Ethnomethodology: Selected readings*. Baltimore: Penguin Books.
- 山森崇義・坂本大介・西尾修一・石黒浩・萩田紀博. (2007). アンドロイドとの「目が合う」条件の検証. HAIシンポジウム 2007 (p. 1B-2). 東京.
(2007年12月25日 受付)
(2008年5月7日 採録)
- [問い合わせ先]
〒192-0981 東京都八王子市片倉町 1404-1
東京工科大学メディア学部
榎本 美香
TEL: 042-637-2111
E-mail: menomoto@media.teu.ac.jp

著者紹介

 えのもと みか
 榎本 美香 [非会員]

2007年千葉大学大学院自然科学研究科博士後期課程修了。2000年ATR知能映像通信研究所研修研究員、2004年千葉大学文学部学術研究支援員、2006年東京農工大学工学府・工学部産学官連携研究員、2007年東京工科大学片柳研究所嘱託研究員を経て、2008年東京工科大学メディア学部助教、博士(学術)。言語心理学の立場から、会話における聞き手の行動・認知メカニズムの解明に取り組む。社会言語科学会企画委員会委員、電子情報通信学会HCG第3種研究会ヴァーバル・ノンヴァーバル・コミュニケーション研究会専門委員、日本心理学会、社会言語科学会、日本認知科学会、人工知能学会各会員。

 なかの ゆきこ
 中野 有紀子 [非会員]

1990年東京大学大学院教育学研究科修士課程修了。同年、日本電信電話株式会社入社。2002年MIT Media Arts & Sciences修士課程修了。2002-2005年(独)科学技術振興機構社会技術研究開発センター専門研究員、2005-2008年東京農工大学大学院工学府特任准教授を経て、2008年4月より成蹊大学理工学部情報科学科准教授。知的で自然なユーザインタフェースの実現に向けて、人との言語・非言語コミュニケーションが可能な会話エージェントの研究に従事。博士(情報理工学)、電子情報通信学会HCG第3種研究会ヴァーバル・ノンヴァーバル・コミュニケーション研究会副委員長、ACL、ACM、人工知能学会、情報処理学会各会員。

Verbal/Nonverbal Behavior Allocation Rules in Social Interactions between Users and Conversational Agents

by

Mika ENOMOTO and Yukiko NAKANO

Abstract :

This paper reports empirical results showing the difference between human-human and human-artifact interactions, and proposes human interaction strategies that are useful in designing human-agents based on human behavioral models. Our empirical study uses a PC operation task as a corpus collection experiment, and reveals human-artifact interaction strategies by analyzing how verbal/nonverbal behaviors are allocated in human-human and human-artifact interactions. First, as basic characteristics in human-human and human-artifact conversations, we found that in human-artifact interactions, the number of utterances and frequency of acknowledgements and other responses are smaller than those in human-human conversations. Then, we propose human-human verbal/nonverbal behavior allocation rules, and examine how these rules are violated in human-artifact interactions, suggesting that these violations are complementary behaviors by the listener that displays understanding of the utterance to the speaker without using a verbal response.

Keywords : human agent, social interaction, non-verbal action, verbal/nonverbal behavior allocation rules, turn-taking system

Contact Address : **Mika ENOMOTO**

School of Media Science, Tokyo University of Technology
 1404-1 Katakura, Hachioji, Tokyo 192-0981, JAPAN
 TEL : 042-637-2111
 E-mail : menomoto@media.teu.ac.jp