

対面対話における発話種別の規則性と身振り

Stylization of Speech Types and Gestures

井上雅史^{1*} 花田里欧子² 古山宣洋¹
INOUE, Masashi¹ HANADA, Ryoko² FURUYAMA, Nobuhiro¹

¹ 国立情報学研究所

¹ National Institute of Informatics

² 京都教育大学

² Kyoto University of Education

Abstract: Many human interaction patterns have been exploited in realizing natural behavior of conversational agents. In addition to the speech styles, non-verbal modalities such as gestures can be utilized. However, it is not clear if non-verbal modalities have stylization that is solid enough. In this paper, we analyze human face-to-face conversation quantitatively and evaluate the usability of co-gesture speech information in human agent interaction.

1 はじめに

対話エージェントをシステムとして設計する際に、例えば質問を投げかけられれば、それに対して応答するのが自然であるといった、人間の発話の定型性に着目し、それらの知識をシステムのふるまいに反映させることができる。本研究では非言語的な活動、特に身振りが、発話に関する規則性をより効果的に使用する上で役立つのかを、人間同士の対話の事例の分析結果をもとに論じる。身振りを伴う発話によって生じる新たな定型性を示唆するとともに、文脈を無視した非言語要素のシステムへの追加が、インタラクションを不自然なものにする危険性を指摘する。

インタラクションにおける非言語的なモダリティーの役割については、エージェントが介在しない電子メールによる人間同士のコミュニケーションにおいても感じ取ることができる。利用可能なメディアがテキストに制限されることで、人々は単なる情報交換を超えた意図の伝達の難しさを感じている。感情伝達が音声を通じてある程度可能となる電話による対話においても、しばしば対面対話では見られないような感情の行き違いが起こっている。これは、我々が通常使用するモダリティーの多様さが、制限されていることが一つの原因と考えられ、対話エージェントにおいても、非言語モダリティーを取り入れることが、より自然なインタラクションを実現する上でも望ましいと考えられる。しかしこれまでのヒューマンエージェントインタラクション

(HAI) 研究では、いかなる状況で非言語モダリティーが効果を発揮するか、またその効果はいかなる機構によって生じているのか、といった観点はからではなく、例えば人間的な身振りを追加提示すれば、その帰結としてインタラクションも自然になるであろう、といった前提が無条件に置かれてしまう場合もありえた。そこで本研究では、実際の人間の対話に基づき、発話種別の規則性を音声言語の観点のみから計量した場合と、非言語コミュニケーションとの関係を考慮に入れて計量した場合を対比させることで、非言語モダリティーにより音声言語のみのインタラクションの表現が如何に変容しうるかを考察する。具体的には、発話内容の規則性を反映する特徴として事前に設定した発話の種類間の遷移パターンを取り上げ、発話と身振りとの共起がその規則性にどのように影響を与えているかを明らかにする。

2 データ

2.1 対象となる対話

本研究では、比較的自由度が高く、かつ目的を持って進められる対話として、カウンセリング場面を取り上げる。現状のHAIシステムにおけるエージェントの役割は、要求される対話の性質に対応して、以下のよう到大別される。1) 実用的な作業を行うための命令を受けつけるもの、2) ユーザと質問応答を行い適切な情報を提示するもの、3) 癒しや楽しみ等の心理的な効果を生むようなインタラクションを実現するもの。

*連絡先: 国立情報学研究所
東京都千代田区一ツ橋2-1-2
E-mail: m-inoue@nii.ac.jp

一番目の対話の典型的な例としては、コマンドを投げかけることで、決まった動作を実行するカーナビゲーションシステムなどが考えられる。これらは、音声応答システムへの置き換えが進められ、実用上の効果を上げている。二番目の定型的な対話については、各種案内や予約受付等での対話が代表的であり、HAI技術の利用の可能性が検討されている。三番目については、ゲーム等のエンターテイメントにおける対話エージェントなど、より人工知能的なふるまいが期待される領域である。

対話の自由度という観点では、どんな話題をどのような順序で話しても良く、なおかつ特定の目的を持たないという条件の、雑談が最も極端な例と考えられる。我々が取り上げるカウンセリング場面は、定型対話と完全な雑談との中間的な性質をもつと考えられる。雑談と異なり、カウンセリング対話では、対話はいくつかの段階を経て進行して行くが、定型対話のようにどのタイミングで何を行うべきかが、事前に決定されているわけではない。クライアントは自身の抱える問題であれば、どのような話題を持ち出しても良い。また、どのように対話を進めても良いが、問題の解決を目指すという目標が設定されており、豊かな文脈を内に含んでいると考えられる。

カウンセリングにおける複数モダリティーの使用については、身振りという非言語的モダリティーを用いることで、言語的な介入に対するクライアントの抵抗感を回避して、カウンセリングを巧みに進行させるといった劇的な事例も報告されている [2]。本研究では、そのように非言語モダリティーが際立った役割を果たす状況を取り上げるのではなく、一見平凡な対話の進行を大局的にとらえる中で現れてくる規則性に焦点を当てる。

2.2 対話の映像

今回の分析の対象となるデータは、2007年から2008年にかけて録画された心理臨床実践場面での対話をビデオに録画したもののひとつである。これらの対話は、カウンセリング技術を学ぶために組織されている自主的な学習会の場において、ボランティアのクライアントが自己の抱える問題を、同じくボランティアのセラピストに相談する場面で、学習会参加者が事後の振り返りを行うために録画されている。本報告で使用するデータはそれらのビデオのうち学術目的での使用を許諾されたものである。カウンセリング対話は、家族療法と呼ばれる心理療法の流派の、特にリフレクティングプロセス [1] という枠組みによって進行する。ただし、これらはカウンセリングの構造についてのみの特徴であり、セラピストが家族療法家であるとは限らな

表 1: 対象となるデータの概要 (長さは分・秒単位)。

Session	Duration	Therapist	Client
1	10.35	男 (中級者)	女
2	27.50		
3	30.32		
4	8.36		

い。リフレクティングプロセスでは、対話の各セッションは、常にカウンセリングに協力する臨床家のグループ (チーム) によって観察されている。この観察結果のコメントが、セラピスト及びクライアントに対して、各セッションの間にチーム内における対話という形式で与えられる。つまり、直接的な働き掛けではなく、インタラクションを観察させることでメッセージを伝達するという構造となっている。このリフレクティングにより、セッションごとの対話の性質の変容が促進される。

今回使用するデータセットの性質は、表 1 にまとめられている。このデータは 4 つの対話 (セッション) からなっており、各セッションの長さはおよそ 20 分であるが、これはカウンセリングの状況やセラピストの裁量によって決定されるもので、事前に制限されているわけではない。ただし、この学習会の期間は半日と設定されており、後日カウンセリングを継続することを前提としていないため、その日のうちに一定の結論に至ることが期待されている。各対話は、セラピストおよびクライアントから見える位置に設置されたビデオカメラおよび外部マイクロフォンによって録画される。前述したように、ビデオカメラは参加者の事後の振り返り学習の目的で設置されており、本研究のために新たに設置されたものではない。従って、参加者のビデオカメラの存在に対する抵抗感は、比較的弱いものと想定されている。

2.3 発話単位

録画されたビデオから、人手によりセラピストの発話区間を同定し、発話種別のラベルに変換する。今回は対話専門家としてのセラピストの行為のみに着目し、クライアントの発話については取り扱わない。これは、HAIにおいては、エージェントのふるまいの設計に着目することに相当する。ラベルを付与する対象となる個々のユニットは、ターン構成要素 (TCU) として切りだされる [7]。ターンは、ある話者が話し始めてから時話者に発話権を譲るために話し終わるまでの発話の全体であり、文・節・句・単語、それらとして完成されていない音の連なり等の、様々な単位やその組み合わせ

表 2: 発話種別 (傾聴・かかわりのレベル)

行為	ラベル
1 開かれた質問	openq
2 閉された質問	closedq
3 はげまし・くり返し	encourage
4 感情の反映	reffeel
5 いいかえ	rephrase
6 意味の反映	refmean
7 その他	other

せにより構成される。TCU を用いた場合、クライアント発話中のセラピストの相槌等の、ターンを構成しない要素はコード化されないため、以降の分析においては「ああ」「うん」といった単純な相槌は、ほとんど対象とされていないことを指摘しておく。

2.4 発話種別

ターンとして切りだされた発話に対し、それぞれが対話内でどのような役割を果たしているかを、発話種別としてコーディングする。発話種別には、臨床心理の実践における対話を把握するために特別に開発された、Ivey らのマイクロカウンセリングのスキーマを用いる [3]。これは、多くの心理臨床の流派に共通する技法を抽出し、臨床実践家がこれを利用して対話を理解し、技能の訓練を行うことを想定して作成されたものである。このスキーマの特徴の一つは、表 2 にあるような質問や反復等の一般的な発話行為に加えて、話題の取り上げ方のレベルや、心理臨床特有の技法に対応した概念のレベルも同時にラベル付けすることである。ただし、ここでは一般性が高く HAI との関連が強いと思われる、表 2 の傾聴・かかわり行為のレベルの発話種別のみを対象として分析を行う。

2.5 身振りの記録単位と種別

録画されたカウンセリング中での、セラピストの両腕または手指の動作のうち、身振りともみなしう動作 (対話を構成する機能を持っていると考えられる動作) が行われた区間を、人手により同定した。身振りの定義は、主として McNeil による [5]。身振りのうち、何かを象徴的に表現する図像的身ぶり、指さしなどの直示的身振りを、表象的、あるいは情報伝達的 (何らかのメッセージ性を持っているもの) な身振りとする。情報伝達的身振りは、一連の動作中での役割により、「準備・ストローク・撤回」に時間的に分節化される。

ここで「準備」動作は、手指がホームポジションから動き出し、「ストローク」に至るまでである。「ストローク」とは、手の動きに強勢が置かれる個所であり、身振りの意味の中心を担う。ただし、ストロークと同様に意味の中心を担うものの、手指が固定された位置や形で持続するものは「保持」の身振りとする。「撤回」は、「ストローク」あるいは「保持」が終了し、手指がホームポジションに戻るまでの動きである。情報伝達的ではない身振りには、上下動等をリズムカルに繰り返す「反復 (ビート)」がある。McNeil による上記のカテゴリに加え、体をさすったり顔の一部に触れたりする「自己接触 (アダプター)」も、情報伝達的でない身振りとして記録した [4]。以上をまとめたものが、表 3 である。

身振りの出現を計測する際に、計数単位を身振りの種類ごとに設定するか、すべての種類の身振りを同一視するかという選択がある。また、情報伝達的身振りにおいては、身振りの始動から終了まで (準備期から撤回期まで) を単位とするか、主要部 (ストローク) のみを問題とするかという選択がある。さらに、左右の手について別々に記録されている身振りについて、それぞれに分析を行うか、両者を統合して、どちらかの手で身振りが行われていればその区間を一律に身振り区間とするか、という選択がある。我々の用いるデータにおいては、身振りの種類やフェーズといった詳細な情報にどのように注目すべきかがまだ明らかではないため、種類を問わず、身振りの生起の有無のみを問題とする。また、フェーズを区分した情報伝達的身振りについては、その発生から終了までを身振り区間とみなす。さらに身振りが片手であるか両手であるかは区別せず、左右いずれかの手が身振りを行っている区間全体を一つの身振りとみなす。

表 3: 身振りコーディングのスキーマ

コミュニケーション的 直示的または図像的 一連の身振りをフェーズに区分 準備, ストローク, 保持, 撤回
非コミュニケーション的 フェーズの区分なし ビート, 自己接触

3 分析

3.1 発話種別の遷移

セラピストが、2.4 で定義された発話種別をどのように使用するかを、発話種別のラベル列から構成された種別間の遷移頻度行列により表現する。カウンセリング対話の冒頭部分の例として第一セッションを、後半部分の例として第三セッションを取り上げ、図1にそれぞれでのセッションでの遷移頻度を図示する。縦軸がある時点でどの発話種別を使用したかを表し、横軸がその次の時点でどの発話種別を使用したかを表す。軸の番号は、発話種別の表の番号と対応している。色の濃い場所が、高い頻度で観察された遷移である。ここでは、クライアントの応答の内容は考慮されていないことに注意が必要である。また、第一セッションは第三セッションより時間が短く、相対的に発話量が少ないため、全体的に遷移頻度図が薄い色となっている。

第一セッションでは、open-q→encourage の連鎖が最も頻度が高く、これは例えば「あなたの抱えている問題についてくわしく話してください」といった問いかけをし、次の発話でクライアントがさらに詳しく話そう、「続けて」などと発話を励ますといった行為の連鎖である。また、次に多い発話を促す行為 encourage→encourage の連鎖では、まずはセラピストが必要な情報を収集するために、クライアントに語らせようとしている様子うかがえる。一方第三セッションでは、前述の二つのパターンは引き続きみられるものの、rephrase→rephrase という、例えば「～ですか」「～ということですね」といった、クライアントの直前の発話を言い換えるような発話の連鎖が増えており、カウンセリングが単に情報を集めるだけの段階から、働きかけてゆく段階へと進行していることが伺える。ただし、第三セッションにおいては様々な遷移パターンが見られ、単純な規則に従って対話が進められているとみなすことはできない。

3.2 身振りを伴う発話の生起

次に、各発話の最中に身振りを行ったか否かにより、発話を「身振りを伴う発話」と「身振りを伴わない発話」に分類する。共起の仕方を問わず、発話と身振りが同時に生成されているオーバーラップ区間が発話内に存在していれば、その発話は身振りを伴うものとした。セッションごとのそれぞれの発話の頻度を、セッション長の違いによらず比較するために、一秒ごとの値として集計したところ、各セッションで 0.0677, 0.0784, 0.0579, 0.0388 となった。つまり、発話頻度そのものは第一セッションから第二セッションにかけて高くなり、その後第三セッションから第四セッションにかけ

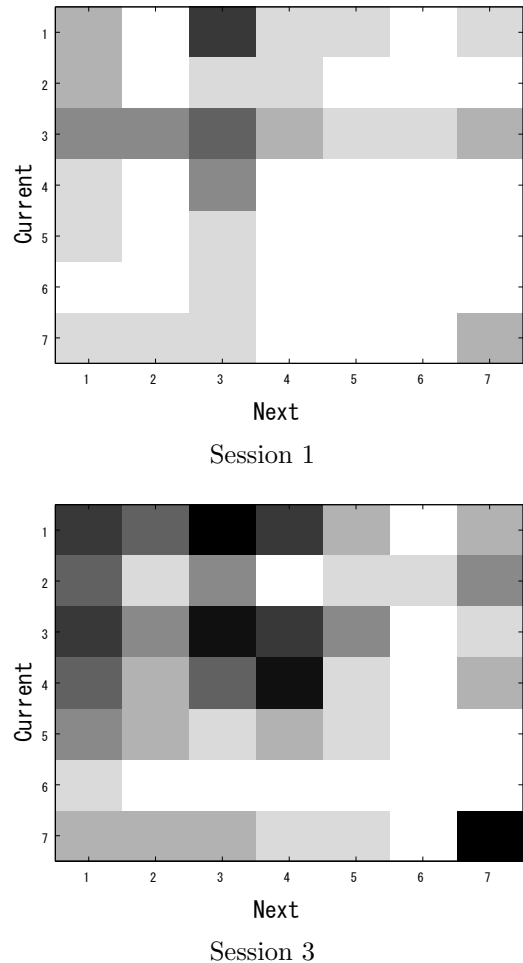


図1: 第一セッション及び第三セッションでの発話種別の遷移頻度。色の濃い部分が頻度が高いことを表す。

て減少している。これはカウンセリングの冒頭においては、セラピストがもっぱら聞き手としてふるまうことを考えると、一見意外なようであるが、クライアントの語りを引き出すために、様々な言葉掛けを行うという「積極的傾聴」が行われているものと理解できる。次に、単位時間当たりの身振りを伴う発話と身振りを伴わない発話との相対的な頻度差に着目する。セッション単位で二種類の発話を秒あたりの頻度で比較したものが、図2である。この図から、セッション1とセッション2では、身振りを伴わない発話が多いのに対し、セッション3とセッション4では、身振りを伴う発話のほうが多くなっていることが分かる。これは、カウンセリングの前半と後半で対話の質が変化しており、後半に向かうに従って身振りの使用が必然となるような状況が生まれていることを示唆している。このことは、カウンセリングと類似した、情報聴取に始まり交渉に進むような構造を持つ対話を行うエージェントを設計する場合には、対話開始直後に発話に伴う身振

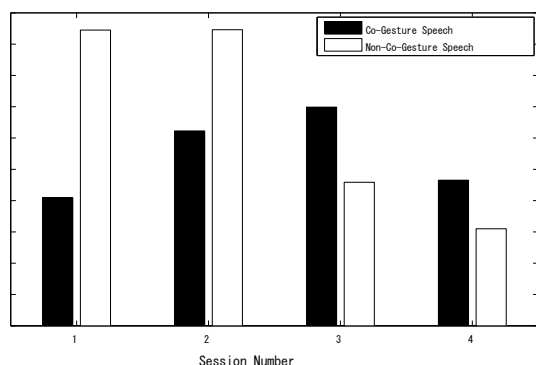


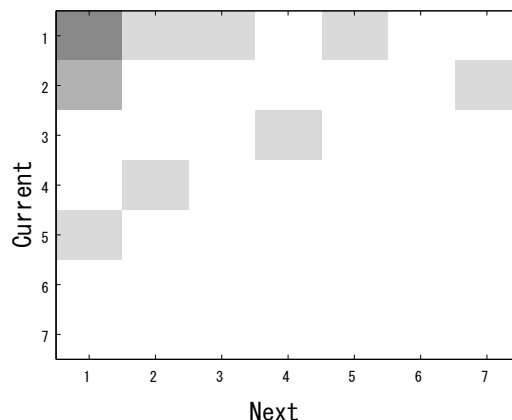
図 2: セッションごとの身振りを伴う発話と伴わない発話の 1 秒あたりの頻度

りを頻繁に使用するような行動規則を導入すると、インタラクションを不自然にしてしまう危険性があることに注意が必要となる。

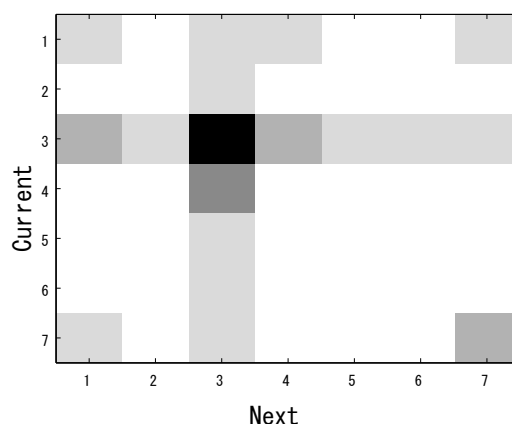
3.3 身振りを伴う発話の遷移

3.1 で行った発話種別の遷移頻度の分析を、身振りを伴うか否かを考慮に入れて、同じデータに対して再度実施して比較することにより、発話のみを考慮した際の規則性が補強されてより鮮明になるのか、あるいは阻害されて判別しづらくなるのかが明らかになる。それにより、非言語モダリティとしての身振りが、対話の規則性に果たす役割が示される。

図 3 内の二つの図は、第一セッションの発話種別の遷移頻度を、身振りとの共起の有無によって二つに分割したものである。上の図では身振りと共に起る発話が、openq の開かれた質問から展開する発話連鎖として取り出されている。下の図では身振りと共に起らない発話が、励ましの連鎖である encourage→encourage を中心に、encourage からの、または encourage への遷移として取り出されているのが分かる。次に、第三セッションの発話種別の遷移頻度を身振りの有無で分割した図 4 内の二つの図を見る。上の図では、第一セッションと同様に openq の開かれた質問に関連する発話連鎖と、第一セッションではほとんど見られなかった rephrase→rephrase を中心とした言い換えに関連する連鎖が、身振りと共に起る発話として取り出されている。下の図では、第一セッションと類似した、励ましに関連した連鎖であるが見出される。このように、発話を身振りとの共起を考慮して分類することによって、発話種別の遷移規則は、より明確に特徴づけられることが分かった。



Session 1: Co-Gesture Speech

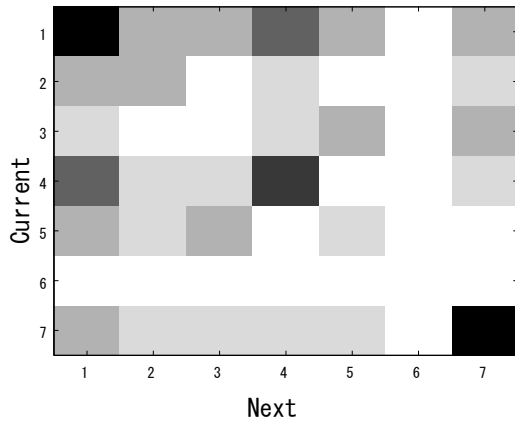


Session 1: Non Co-Gesture Speech

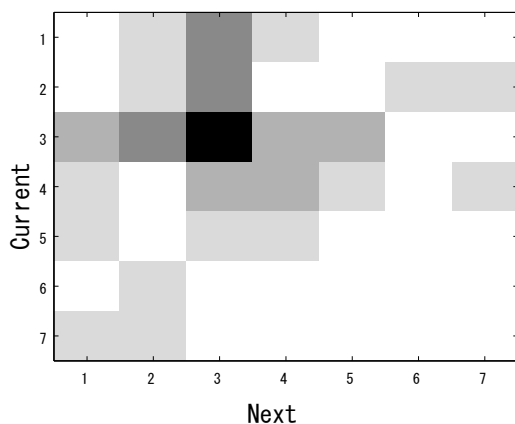
図 3: 第一セッションでの身振りを伴うまたは伴わない発話の発話種別の遷移頻度。色の濃い部分が頻度が高いことを表す。

4 考察

これまで見てきたように、身振りとの共起を考慮に入れることで、発話連鎖規則の定型性をより明瞭に捉えることができる。このことを利用すれば、対話エージェントに、より効果的な行動規則を埋め込むことができる。例えば、相手の発言を受けて行う「！と言うと...」といった発話では、「具体的に言うというのでしょうか」という意図の、詳細に興味のある「質問」がなされているのか、「続きを話してください」という意図で発話を擬似的に反復しての「励まし」がおこなわれているのか、はっきりしないことがあり、エージェントが利用した場合にユーザを混乱させる可能性がある。本研究で得られた知見に従えば、このようなときに発話に身振りを伴わせるか否かによって、その発話が質問であるのか励ましであるのかを、明確にすることができる。



Session 3: Co-Gesture Speech



Session 3: Non Co-Gesture Speech

図 4: 第三セッションでの身振りを伴うまたは伴わない発話の発話種別の遷移頻度。色の濃い部分が頻度が高いことを表す。

最後に、人間同士の対面対話に規則性を見出そうとする他の研究との関連について述べる。本研究と同様のカウンセリング場面での対話を取り上げた研究として、模擬カウンセリングを対象とした長岡らの研究がある [6]。本研究が個々の発話の遷移という比較的短時間のパターンに着目しているのに対し、長岡らはカウンセリング全体を通じた身体と同調傾向の変化を取り上げている。また、本研究で取り上げた発話と身振りに加え、視線も加えた共起関係に着目した試みとして、澤本らの医者と患者の模擬面接を対象とした研究がある [8]。ただし、我々が規定したような、個々の発話の役割については設定されていない。

5 おわりに

本研究では、定型性を持つ言語的な情報に基づくインタラクションの表現が、非言語情報を考慮に入れる

ことによってどのように影響されるのかを、発話種別の遷移パターンの規則性の変化として把握することを試みた。特に発話種別と身振りとの関係に着目し、身振りと共に発話と共起しない発話に特徴的な発話種別の連鎖規則を見出すことができた。このことは、非言語モダリティーを適切に用いることで、HAIシステムをより自然なものにしようことを示唆している。同時に、身振りと共に発話のカウンセリング対話の初期にはあまり出現しないことなど、文脈を考慮しないモダリティー追加の危険性も明らかになった。以上の論議は、カウンセリングの一事例の分析に基づいている。今後はより多くの事例を検討し、上記の論議の妥当性を検証することが課題である。

参考文献

- [1] T. Andersen. *The reflecting team-dialogues and dialogues about the dialogues*. W.W. Norton, New York, 1991.
- [2] R. Hanada, N. Furuyama, and M. Inoue. Speech-gesture mismatch and how it changes in a problem described by a client. In *International Society for Gesture Studies, Third International Conference*, pages 5–6, Evanston, Illinois, June 18–21 2007.
- [3] A. E. Ivey and M. B. Ivey. *Intentional Interviewing and Counseling*. Brooks/Cole, 5th edition, 2002.
- [4] M. Kipp. *Gesture generation by imitation: from human behavior to computer character animation*. Dissertation.com, 2003.
- [5] D. McNeill. *Hand and mind*. The University of Chicago Press, 1992.
- [6] C. Nagaoka and M. Komori. Body movement synchrony in psychotherapeutic counseling: A study using the video-based quantification method. *IE-ICE Trans. Inf. & Syst.*, E91-D(6):1634–1640, June 2008.
- [7] 榎本美香. 会話・対話談話研究のための分析単位? ターン構成単位 (TCU). *人口知能学会誌*, 23(2):265–2709, March 2007.
- [8] 澤本祐一, 平野靖, 梶田将司, 間瀬健二, 勝山貴美子, 神山祐一, 山内一信. 医療面接に特徴的なインタラクションパターンの抽出方法. In *医療情報学*, volume 27(Suppl.), pages 931–934, 2007.