非言語行動の出現パターンによる会話構造抽出

中田　篤志[†]　　角　康之[†a)]　　西田　豊明[†]

Sequential Pattern Analysis of Nonverbal Behaviors in Multiparty Conversation

Atsushi NAKATA[†], Yasuyuki SUMI[†a)], and Toyoaki NISHIDA[†]

あらまし　我々は会話中に，視線，指差し，うなずきといった様々な非言語行動を行っている．それらの非言語行動は会話の制御に使われていると考えられ，その出現パターンには一定の構造がある．本研究ではこれを会話構造と呼ぶ．本研究では，会話参加者らによる非言語行動出現の時間構造を N-gram で表現し，会話記録のデータから会話構造を自動抽出するインタラクションマイニングの手法を提案する．そして，提案手法を用いてポスター発表会話とポスター環境自由会話という 2 種類の会話状況における会話構造の自動抽出を試みた．その結果，発話者は非発話者より指差しが多い，とか，うなずきの後に相槌を行うことが多いといった会話構造は二つの会話状況に共通して見られる一方で，沈黙の後に発話を続けるのは元の発話者であるという会話構造はポスター発表会話特有のものであるといったことを確認することができた．

キーワード　多人数会話分析，非言語行動，インタラクションマイニング，会話状況認識

1. まえがき

我々は会話を行う際，言語による情報だけでなく，視線・指差し・うなずき・相槌といった多くの非言語行動によって会話相手に自分の意思を伝えている．そして，それらの非言語行動は出現パターンにはある程度の規則性があると考えられる．例えば「会議の議長は視線を遷移させやすい」,「発話をする際には，同時にジェスチャを伴う場合が多い」,「現発話者から視線を向けられた人物はうなずきを返す」といったものが考えられる．会話状況や会話参加者に依存して発生するような，このような非言語行動の出現パターンを，本研究では会話構造と呼ぶ．

我々がインタラクションを行う環境には，立ち話やミーティング，共同作業中の会話など様々なものがあり，それによって会話の場所・人数・参加者の会話内での役割・会話の目的といった属性が存在している．異なる会話環境でも共通する会話構造がある一方で，それぞれの会話環境に依存する会話構造もあると考えられる．例えば「新しく話し始める参加者は，前に話していた話者から視線を向けられている場合が多い」といった会話構造は，立ち話やミーティングでは成り立つと考えられる一方，作業対象に視線が集まりやすい共同作業では成り立たない場合が多いと考えられる．

筆者らは様々な非言語行動を伴う会話データを大量に蓄積することで，異なる会話環境における会話構造の共通性や差異を比較したいと考えている．異なる会話環境において共通する会話構造を明らかにできれば，複数の環境において自然に振る舞うロボットの設計に有用であろうし，会話環境に依存した会話構造を明らかにすれば，その知識をミーティング記録データの場面識別に応用することができるであろう．

そこで，本研究ではデータマイニングの手法を用いることで，非言語行動の出現パターンとして表現される会話構造を抽出するための枠組みであるインタラクションマイニングを提案する．本手法の有用性を検証するため，ポスター発表会話とポスター環境自由会話という二つの会話環境で会話データを収録し，それらの会話データをインタラクションマイニングを用いて分析する．まず，それぞれの環境における会話構造を抽出し，日常の会話でも用いられているような会話構造が抽出されているかを確認する．その上で，ポスター発表会話とポスター環境自由会話という異なる会話環境において抽出される会話構造の相違を比較する．

[†] 京都大学大学院情報学研究科，京都市
Guraduate School of Informatics, Kyoto University, Kyoto-shi, 606–8501 Japan
a) E-mail: sumi@acm.org

2. 関連研究と本研究の位置付け

会話的インタラクションにおける構造を明らかにしようとする試みは会話分析研究として数多く行われてきた [1]〜[4]．それらは，分析者が視線，ジェスチャ，うなずきといった特定の非言語行動に着目し，それらと発話との時間的共起関係を分析して，発話交替や発話意図を推定するモデルの構築を目指すものである．そこでは，非言語行動の出現パターンに関する仮説が先にあり，それをデータ分析により検証するという研究の方式がとられる．本研究は，観測し得る非言語情報を大量かつ多様に収集し，そこから非言語情報の共起出現パターンを網羅的に数え上げ，そこからボトムアップ的に特徴的な会話構造を抽出することを目指しており，上記のような会話分析に対して，分析の観点を与えるものであると考える．

多数のカメラやセンサを用いて大量に会話データを収録し，インタラクションの分析を試みている研究として AMI [5] や VACE [6] と呼ばれる研究プロジェクトは筆者らと立場が近い．しかし，彼らの研究の焦点は非言語行動の自動検出や談話構造の解釈であり，会話の構造理解については従来の仮説検証型の手法が用いられている．また，対象とする会話環境はミーティング形式に特化されている．それに対し，大塚ら [8] は，会話参加者の視線パターン，頭部方向，及び発話の有無に基づいて会話構造を確率的に推定する枠組みを提案している．数理的なモデルにより会話構造を理解しようという目標を本研究と共有する．しかし，大塚らの研究では，対面着座式の 4 人会話に特化し，その中で会話構造を推定するモデルの精度向上を目的としている．それに対して本研究は，様々な会話モードを含む会話データから，会話構造出現の傾向や分析の観点を見出すための枠組みをつくることを目的としている．

非言語行動に関する計測データから会話構造の理解を試みた先行研究としては，森田らの研究 [7] が挙げられる．そこでは，赤外線 ID センサにより自動付与された滞在・注視といった非言語行動の時間的出現パターンの抽出を試みた．本研究はその拡張として，計測する非言語行動データを増やし，更に，非言語行動出現パターンを N-gram 表現で定式化して会話状況ごとの特徴抽出を試みる．

3. インタラクションマイニング

インタラクションマイニングは，データマイニングの手法を利用し，非言語行動の時間的出現パターンに着目してモデル化を行い，会話状況に依存した会話特徴抽出を行う手法である．ここでは，インタラクションマイニングの考え方と具体的な手順を説明する．

3.1 インタラクションマイニングの基本的考え方

インタラクションマイニングでは，以下の三つの考え方に基づいて会話的インタラクションをモデル化する．

(1) 会話中に，異なる非言語行動が前後して出現するパターンに着目する．

(2) 非言語行動の出現回数を数え上げる際に，対等な関係の会話参加者が関与する非言語行動は，同一のものとして数える．

(3) ある会話状況において，非言語行動の生起回数に偏りがある場合，それは特徴的会話構造である可能性が高いと考える．

(1) は従来の会話分析で用いられてきた考え方である．(2) についても同様で，会話参加者 A から D が対等な立場で参加している場合，多くの会話分析では「現話者 A が次話者 B を見ていた」という視線生起の回数と，「現話者 B が次話者 A を見ていた」という回数や，別の収録データで「現話者 C が次話者 D を見ていた」という回数を同一の現象として合算して数え上げを行っている．(3) については，例えば榎本ら [2] は「発話交代が起こる場面」に着目し，前後の非言語行動の生起回数を数え上げ，それらの生起回数や発生時間に偏りが見られた場合に会話構造が存在すると結論づけている．インタラクションマイニングは，前後する非言語行動の生起回数が偏っている部分を網羅的に抽出することで，特定の会話場面に限定することなく会話構造を抽出することを目指す．

非言語行動の時間的出現パターンを表現する方法として，本論文では N-gram を利用する．N-gram は自然言語処理研究でよく使われており，ある文字列の中で N 個の文字（あるいは単語）の組合せがどの程度出現するかを数え上げるモデルである．例えば $N-1$ 個の単語を見て次の単語を予測するモデルに使われる．本研究では，非言語情報が共起した状態（後述するインタラクションステート）を単語と見立てて N-gram を適用することで，非言語情報の出現の時間的変化を表現する．

本研究で N-gram を利用する特徴として，会話参加者個人ごとの各非言語情報を単語とするのではなく，会話参加者全員の非言語情報の共起状態を単語とする．つまり，会話における非言語情報は，個人個人で独立に生起するのではなく，会話相手の反応と呼応しながら生起すると考えられ，その時間的変化こそが会話状況を特徴づける指標になり得ると考えている．

3.2 インタラクションステート

会話的インタラクションにおける非言語情報の出現パターンを表現する要素（N-gram における単語）として，インタラクションステート（以下ステートとも記述）という概念について説明する．

インタラクションステートとは，例えば「全員で一つの展示物を見ていて，1 人がその展示物を指差している」といった会話の一場面を表現したものである．ステートは，分析対象として計測された非言語行動それぞれについての状態（例えば，発話をしているか否か等）の組合せで表現される．つまり，本論文では，会話データとして 3 人会話を選び，計測対象となる非言語行動として，発話・相槌・うなずき・指差し・視線を利用する．発話・相槌・うなずきについては，それらの発生の有無だけの 2 状態で表現されるのに対し，指差しと視線については，発生有無に加えて，それが向けられた対象（対象物としてのポスターや本人以外の会話参加者）の数だけ状態がある．インタラクションステートは，これらの非言語行動が共起した状態を表すので，起こり得るステートの種類としては，3 人についての上記の状態組合せ数だけ存在する．

インタラクションステートの例を図 1 に示す．この例では，会話参加者を○，会話参照物としてのポスターを□で表現し，発話・指差し・視線注視の有無と方向を記号で表している．S の例は，3 人が向き合いながら会話をしている一場面で，1 人が発話し，あとの 2 人がその発話者に視線を向けている，という状況である．発話者は片方の聞き手に視線を向けている．T の例は，ポスターの横に立ちながら 3 人が会話している一場面を示し，1 人がポスターを指差しながら発話している．聞き手の 1 人は発話者に視線を向けており，もう 1 人はポスターに視線を向けている．これらの例から分かるとおり，本論文で扱うようなポスター会話に限定されず，参照物のない会話状況も表現できる．また，3 人会話に限らず，2 人対話や 4 人以上の会話でもインタラクションステートは表現できるし，あるいは，1 人が複数の対象物を眺めながら歩き回るような状況も同様に表現することが可能である．

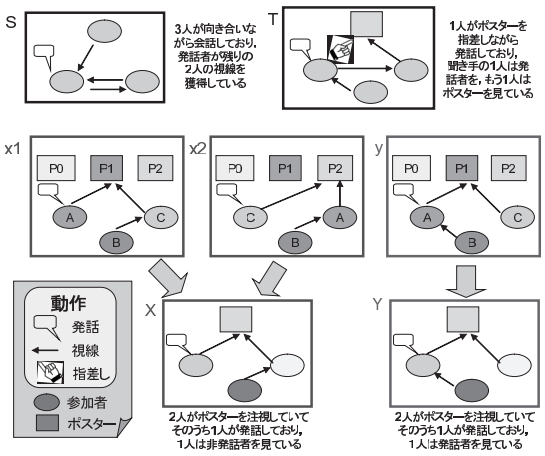

図 1　インタラクションステートの例
Fig. 1　Examples of interaction state.

ここで，会話参加者や対象物の入換えによるステートの同一判定について説明する．インタラクションステートでは以下の条件をすべて満たすとき，二つの会話状態を同じステートとして扱う．

- 会話参加者または非言語行動の対象物である a, b が，分析者によって同一とみなされる
- a, b を入れ換えたときに，各参加者がとっている非言語行動の発生の有無及び対象が同一である

ステートの同一判定についての例を図 1 の X と Y に挙げる．ここでは分析対象は視線・発話の二つの非言語行動とする．また，すべての会話参加者，及び視線の対象となるポスターの立場は同一とみなされているとする．

ここで，ステート x1 とステート x2 は，参加者 A と C を入れ換え，ポスター P1 と P2 を入れ換えることで同じ状態となるため，同じステート X として扱う．一方，ステート y はステート x1 と A, C の状態は同じであるものの，B が発話者である A を見ているか，非発話者である C を見ているかという点が異なるため，異なるステート Y として扱う．

なお，図 1 の例はあくまで全員が入換え可能である場合の例である．仮にこれがポスター発表の場であり，分析者が「A は発表者であるため B, C とは同一とみなさない」と考えた場合や，ポスター P1 に特に重要な情報があり他のポスターと同列にはみなせない場合には，x1 と x2 は別のステートとして扱われることになる．

3.3 インタラクションマイニングの手順

ここではインタラクションマイニングの手順を説明する．

3.3.1 多人数会話のデザインと収録

分析の基礎データとなる多人数会話の環境を設計し，その環境下でなされた会話データを収録する．大量のデータに対してラベリングを行うことを考えると，様々なセンサを用いて非言語行動を収録することが重要である．

3.3.2 非言語行動のラベリング

得られた会話データから，非言語行動がいつ始まり，いつ終わり，何を対象としているかについて情報を付与する．本論文ではこの行為をラベリング，作成された付加情報をラベルと呼ぶ．

ラベリングは多くの会話分析で行われていることであるが，インタラクションマイニングでは多種の非言語行動を大量にラベリングする必要がある．そこで，インタラクションマイニングの対象とするデータの収録時にはモーションキャプチャや視線計測装置といったセンサを用いることで，一部の非言語行動のラベリングを補助している．

3.3.3 分析対象とするラベルの選択・前処理

作成したラベルから分析対象とするラベルを選択する．細かすぎるラベルはインタラクションステートの変化を過度に引き起こす原因となるため，それを除去する．これは，通常のデータマイニングで行うデータの前処理に相当する．この手続きは，後述する抽出結果の検証を踏まえながら再帰的に行う．

3.3.4 人・対象物の同一性決定

3.2 で述べたように，参加者や対象物の役割に差がある場合，それらを別個として扱った方がよい場合がある．そこで，インタラクションマイニングを行う際にはそれぞれの参加者を同一とみなすかどうか，対象物を同一とみなすかどうかを分析者があらかじめ定義しておく．

3.3.5 インタラクションステートへの変換

3.2 で述べたインタラクションステートの考え方をもとに，分析対象となっている非言語行動のラベルからステートの列を作成する．ステートは注目した非言語行動の状態を記述したものであるため，いずれかの非言語行動の発生の有無・対象が変わるたびにステートが変化する．この変換の例を図 2 に示す．

なお，この時点では参加者やポスターは別個に扱い，参加者・ポスターの情報を残しておく．

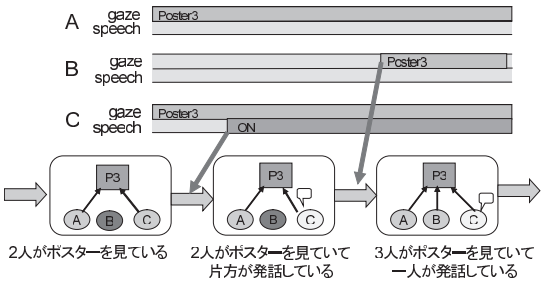

図 2 ラベルデータからインタラクションステート列への変換例

Fig. 2 Converting labels into a sequence of interaction states.

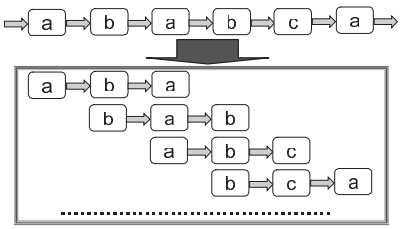

図 3 ステート列の分割

Fig. 3 Extraction of sequence of interaction states.

3.3.6 枝要素の生成

得られたステート列を一定の長さで切り分け，会話参加者・ポスターの同一性について判断する．これは，N-gram モデルで表現される単語列に相当する．

ここで生成されるステート列を枝要素と呼ぶ．以下，枝要素の生成過程を説明する．まず各ステートを起点にステート列を一定の長さで取り出す．この際，切り分ける長さは特徴的構造が得られる範囲で十分に大きな値を設定する．ここで，切り出す長さを 3 とした場合（つまり 3-gram）の例を図 3 に示す．この長さを 1 にすれば（つまり 1-gram），ステートごとの出現頻度を数え上げることになる．

次に，ステートの同一性について，「枝要素の中で同一でないと認定された参加者や対象物は，以後のステートでも同一でないとして扱う」という新たな条件を追加した上で再認定を行う．例を図 4 に挙げる．ここでは非言語行動は発話と視線のみを対象とし，枝要素の長さを 3 としている．α の枝要素において，1 番目のステートで A は発話を行っている．そのため，2 番目のステートで A が沈黙した後でも，A は B, C と同一とはみなさない．よって，3 番目に C が話し始めたときには，全員が同一でないとして扱う．一方で，β の枝要素では 1 番目で A, B の間に非言語行動の差

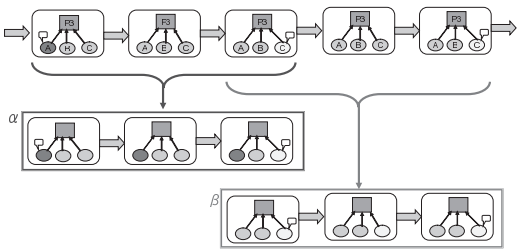

図 4　枝要素内における参加者の同一性認定例
Fig. 4　Unification of subjects in state sequence.

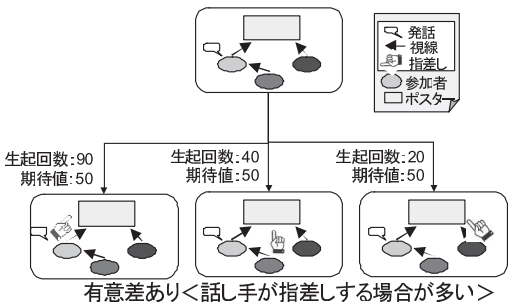

図 5　χ^2 検定による構造抽出
Fig. 5　Structure extraction by χ^2 test.

がないため，これらは同一のものとして扱う．このように，人・ポスターの同一性は同じステートを対象としていても枝要素によって異なる場合がある．

このようにしている理由は，「以前発話をしていた」という情報が会話構造において重要な意味をもつ場合があるからである．例えば人が交互に話すのが一般的である会話では「以前発話していた人物は発話終了後に再度発話する回数が有意に少ない」という特徴的会話構造が存在する可能性がある．これと同様に，「以前ある非言語行動を行っていた」という事実が会話構造に現れる場合が多くの会話において存在し，このような会話構造を抽出するために前述したような参加者の同一性判定を行っている．

3. 3. 7　木構造の生成

得られた枝要素を木構造の形にまとめる．これにより，生起したステートと，それを起点としたステートの遷移，及びそれぞれの遷移の生起回数を得ることができる．

3. 3. 8　χ^2 検定による特徴的構造抽出

木構造を生成した後，「インタラクションステートの遷移において，次に行われる同一の非言語行動について偏りは存在しない」という帰無仮説をもとに χ^2 検定を行い，仮説が棄却される木構造を抽出する．

具体例を図 5 に示す．この図ではステートが変化した場合のうち，変化がポスターへの指差しによるものであった場合のみを示している．

最初の状態では全員が指差しを行っていない．ここで前記の帰無仮説を採用すると，指差し行為について言えば，A，B，C のだれもが次の行動として指差しを開始する可能性があり，その確率は 1/3 で等しい．すなわち，図 5 の 2 段目のステートへの遷移が同じ回数ずつ起きるということになる．しかし，実際の生起回数は，それぞれ 90 回，40 回，20 回のように偏った場合，χ^2 検定において帰無仮説が棄却され，この木構造を特徴的な会話構造として抽出することができる．

これにより，前項で生成された非常に大きな木から，特徴的な会話構造と思われる部分のみを体系的に抽出することができる．

3. 3. 9　抽出結果の検証

χ^2 検定によって抽出された会話構造の中には，自動ラベリングにおいて混入した誤認識や，特定の人物・収録データに過度に依存した会話構造なども含まれ，必ずしも注目に値する特徴的構造であるとは限らない．一方でもし特徴的構造があっても，他の非言語行動によるステート変化にまぎれてしまい十分に偏りが見られない場合もあり得る．

以上のような理由から，通常のデータマイニングと同様に χ^2 検定によって得られた結果は分析者が確認し，採用すべき特徴的な会話構造であるかを確認する必要がある．この結果を踏まえて，まだ十分に会話構造が得られていないと考えた場合は **3. 3. 3** に戻り，分析対象とする非言語行動を選び直したり，ラベルの処理を行うといったことを繰り返し行う必要がある．

この一連の作業を行うために，筆者らは，N-gram の N の値を変えたり，人・参照物の同一判定を行ったりしながら，インタラクティブに木構造を抽出・評価するための GUI（グラフィカルユーザインタフェース）を試作し，利用している．

4.　データ収録とラベリング

本研究ではインタラクションマイニングによる会話分析の対象として，ポスター発表会話とポスター環境自由会話という二つの会話を対象にした．それぞれの収録の様子を図 6 に示す．ここでは二つの会話環境の設計について説明し，分析対象とする非言語行動とラベリング手法について説明する．

図 6　収録した会話環境
Fig. 6　Recorded conversation environments.

4.1　参照物が存在する 3 人会話

インタラクションマイニング自体は，対象物としてのポスターが存在する 3 人会話の分析に限定されたものではなく，2 人や 4 人以上の会話であるとか，ポスターなどの対象物が存在しない対面着座式の会話にも適用可能である．本論文で，参照物としてのポスターが存在する 3 人会話に着目したのは，一見，似通った環境に見える 3 人会話の中に，様々な会話状況が発生することを示したいと考えたからである．

本論文でデータとして利用したポスター発表会話とポスター環境自由会話は，どちらも会話参加者が 3 人で，参照物としてポスターが存在する，という意味では，大変似通った会話環境といえる．しかし，ポスター発表会話の場合は立ち位置が比較的固定され，また，会話参加者の間に発表者と聞き手という明らかな役割分担がある．したがって，3 人の発話量には大きな偏りがあり，ポスターに対する視線や指差しが頻出する，といった特徴があると期待される．また，そういった状況の中にも，発表者が一方的に説明しているモードや，聞き手の質問をきっかけとしてよりインタラクティブな質疑応答モードが含まれ，3 人会話としての多様性を含んでいると期待される．

一方，ポスター環境自由会話では，複数（6 枚）のポスターで囲まれた空間を自由に歩き回りながら会話をするため，必ずしも 3 人が同一の会話場に参加しているとは限らず，各自が黙りながら別々のポスターを見ている状態もあれば，ある参加者の一言で 3 人が一箇所に集まり同一のポスターを見ながら（あるいはポスターは関係なしに）会話を始める，というような動的な状態遷移も生じる．つまり，限られた会話環境にもかかわらず，その中には，発話交替，発話とジェスチャの共起，視線配分といったミクロな会話構造や，会話場の発生と移動といったマクロな会話構造について，様々な種類を観測できることが期待される．

そして，そういった様々な会話状況に対応した特徴抽出の手段として，インタラクションマイニングが有効であることを示すのが本論文の目的である．

4.2　ポスター発表会話

ポスター発表会話は，あまり専門的な議論にならずに，聞き手 2 人が自由に質問できる環境を作ることに配慮した．具体的には，発表者には，非専門家に対して話すようにポスターの準備と話題の選択をお願いした．また，身近な人同士だけで通じるような細かい議論を避けるために，初対面あるいは特に親しくない人同士で発表者と聞き手の 3 人の組を作った．データ収録の参加者は，筆者らの研究室外から集め，生物，物理，工学を背景とした程良く内容が分散した分野の大学院生を集めることができた．したがって，聞き手の興味や理解を保つことができ，自然に質問応答がなされる会話を収録することができた．

発表者には 10 分程度の発表をお願いし，聞き手には自由に質問をしてもらうよう教示した．データ収録は会話参加者を入れ換えながら 8 回行い，そのうち計測データの欠損が少なく，かつ，活発に会話が行われた 4 セッションを選んで分析対象とした．それら 4 セッションとも，ほぼ 20〜25 分程度の会話データとなった．

4.3　ポスター環境自由会話

ポスター環境自由会話は，より自由かつ探索的に話題を変えながら会話をしてもらうことを期待し，筆者らが所属する研究室内のよく知った学生同士で 3 人の組を作って収録した．話題環境として，洛中洛外図と呼ばれる室町時代の京都市内を描いた屏風絵を利用した．これは，京都市内で生活をしている学生が探索的に身近な話題を見つけやすくするためである．参加者は，歴史や美術の知識はあまりもっていなかったため，実験者である筆者らがあらかじめ屏風絵の中に馴染みの地名や寺社名（「百万遍」や「金閣寺」など）を付箋で貼り付けて，屏風図を見るためのとっかかりを用意した．また，話題のヒントとして，「現在も残る建物を探してみてください」，「描かれた時代を推定してみてください」といった教示を与えるなどの工夫をした．

データ収録は会話参加者を入れ換えながら 3 回行い，そのうち計測データの欠損が少ない 2 セッションを選んで分析対象とした．その結果，1 セッション当り 20 分前後，自由に話題を変えながらの会話データを収録することができた．

4.4 対象とする非言語行動

本研究では既存の会話分析研究でも着目されてきた非言語行動を分析の対象とし，インタラクションマイニングの手法でも従来から議論されているような会話構造の抽出が行えるかを検証する．そのため，発話・相槌・うなずき・ジェスチャ・視線を分析対象とした．

ただし，ジェスチャに関しては多種多様なものが考えられ，すべてのジェスチャを定量的に扱うのは困難である．そこで本研究では，ポスターが存在する環境で特に重要な役割をもち，かつ対象物や意味を特定するのが容易である指差し行為を扱うこととした．また，うなずきについては多くの会話分析では聞き手のみが行うものとして扱われるが，本研究ではメイナードの研究 [4] で議論されているように，発話者によるうなずきも含めて分析対象とした．

4.5 ラベリング

ここでは，分析対象とする非言語行動のラベリングについて説明する．

4.5.1 発話及び相槌

発話区間の認定，及び相槌と通常発話の認定については手作業で行った．まず，無線マイクによって収録した発話音声をもとに，日本語話し言葉コーパス [9] を基準に発話区間の認定と発話内容の書き起こしを行った．次に，吉田らの研究 [10] を参考にして通常発話と相槌の分離を行った．

4.5.2 指　差　し

指差しの認定は，Kendon [11] が提案している「ジェスチャー単位」及び「ジェスチャー節」を参考にして手作業で行った．

4.5.3 視　　線

視線情報は会話参加者が装着したモーションキャプチャのデータ，及び視線計測装置のデータから，自動的にラベルを生成した．視線ラベルの生成にあたっては，まず視線計測データと頭のモーションデータから視線ベクトルを計算し，他の人の頭部をモデル化した球体や，ポスターをモデル化した長方形との衝突判定を行った．更にセンサからデータが取得できない場合があることを考慮し，250 ms 以下の短いラベルに対して補間を行った．各種しきい値などは，先行研究 [12] の際に最も精度よく取れたものを利用している．

4.5.4 う な ず き

うなずきは加速度センサによって取得した三次元加速度情報をもとに自動ラベリングを行ったものを利用した．抽出はまずうなずきの候補となる動作を抽出した後，体全体の動作によって生じるノイズを除去するという手法で行った．うなずき候補の抽出は，加速度センサの Y 軸方向と Z 軸方向の加速度の絶対値に対して 600 ms の区間で分散をとり，その中央値と偏差値から定めたしきい値よりも分散が大きい区間を抽出した．ノイズとしては顔方向を変える動作，上体を傾ける動作，体全体が揺れる動作を扱い，参加者の頭・背中・腰につけた加速度センサで抽出した．以上の三つの動作をノイズとしてうなずき候補から除去し，残ったものをうなずきラベルと認定した [13]．

4.6 収録データの性質

N-gram を用いた会話構造抽出に入る前に，収録した会話がどのようなデータになったかを概観しておく．

ポスター発表会話は，予想どおり，発表者と聞き手で発話量が大きく異なり，データ分析に利用した四つのセッションに共通して，発表者の発話総時間は 12〜16 分程度であるのに対し，聞き手の発話総時間は 1〜2 分程度だった．また，指差しジェスチャについてはその差は更に顕著で，発表者はポスターへの指差しを 80〜120 回程度したのに対し，聞き手は 0〜7 回だった．視線については，発表者，聞き手ともにポスターに対して向けられている時間が多く，聞き手は発表者に対してほどほどに視線配分を行い，もう一方の聞き手に対して視線配分をすることは非常に少ない，という状況だった．なお，発話量，指差し，視線は，発表者による説明の時間帯と，質問応答の時間帯では大きく様子が異なり，そのことは，ステートの発現状況の偏りからも確認することができた．

ポスター環境自由会話では，一つのセッションでは 3 人がほぼ同等に発話している（約 3〜6 分ずつ）のに対し，もう一つのセッションでは 1 人が聞き手に回っている（2 人が約 8 分，5 分なのに対し，残りの 1 人は 1 分少々）状況だった．その 1 人については，二つのセッションに参加した 6 人のうちで有意に相槌の回数が多いことや，ポスターへの指差しが少ないことも確認できた．それ以外の 5 人については 20〜40 回程度の指差しを行っており，六つのポスターに対して満遍なく指差しが分散していることから，話題が適当に遷移していることを確認することができた．視線も満遍なく様々な対象に配分されていたことが確認された．ただし，各々のポスターに向けられた時間よりも，自分以外の会話参加者に対して視線を向けた総時間が長めであることから，ポスター発表会話に比べて，より自由に身体位置関係を変えながら，会話を行っていた

ことが確認できた．

5. 各会話環境において抽出された会話構造

前章で作成したラベルデータをもとにインタラクションマイニングを行った．ここでは，それぞれの会話環境で得られた会話構造とそれらがどのような形で木構造に現れるかについて説明する．

5.1 ポスター発表会話で得られた会話構造

ポスター発表会話で得られたラベルデータをもとに，ラベルの前処理と抽出結果の検証を繰り返しながらインタラクションマイニングに取り組んだ結果，以下のような会話構造を得ることができた．

(1-1) 発表者は被発表者より発話を始めやすく，終了しやすい
(1-2) 発表者は被発表者より指差しを行いやすい
(1-3) 発表者は被発表者よりポスターから視線を外すことが多い
(1-4) 発表者は被発表者よりうなずきを行いやすい
(1-5) うなずきを行った人物は他の参加者よりも相槌を行いやすい
(1-6) 発表者の視線によって，被発表者の連続発話に差異が生じる

本論文では，これらのうち (1-3) と (1-6) を取り上げ，会話構造が抽出された木構造にどのように現れるかについて説明する．

5.1.1 ポスターからの視線遷移に関する会話構造

「発表者は被発表者よりポスターから視線を外すことが多い」という結果は，図 7 のような木構造に現れている．木構造のはじめの状態から視線を外す回数は,「全員の非言語行動の遷移に偏りがない」という帰無仮説に基づくと，発表者が 1 人に対し被発表者が 2 人であるためそれぞれ 1/3，2/3 となる．これに対し，実際の生起回数は発表者の遷移回数が 41 回，被発表者 2 人の遷移回数の合計が 26 回であったため，この木構造に有意差があるとして抽出された．

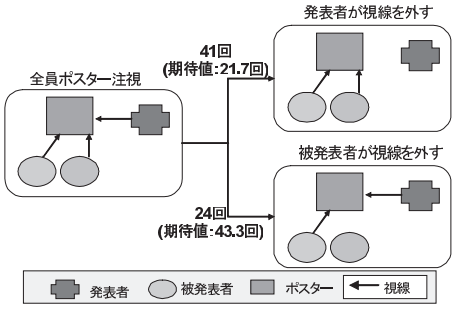

図 7 ポスター発表会話における視線遷移の会話構造
Fig. 7 Conversation structure about gaze transition in poster presentation.

また，このポスターから視線を外す場合の生起回数の偏りに関する構造は，発表者が発話を行っている場合や指差しをしている場合にも見られた．そのため，この視線に関する構造は非言語行動よりも発表者という参加者の役割に起因すると考えられる．

この構造に関して実際に動画や音声から検証したところ，ポスター発表会話では発表者が被発表者の様子を見るために頻繁に被発表者の方を見るのに対し，被発表者はポスターからあまり視線を動かしていない様子が見られた．このような参加者の振舞いの差が会話構造として抽出されたのではないかと考えている．

5.1.2 発表者の視線による被発表者の発話差異

被発表者が発話を終了させた後，発表者が被発表者を注視している場合には，再び被発表者が発話する回数が他のいずれのと比べても有意に多いことが分かった．その一方で，発表者がポスターを注視している場合，発話終了後に再び被発表者が発話する回数はもう一人の被発表者とは有意差が確認されたものの，発表者が発話する回数とは有意差が見られなかった．前者の会話構造を X，後者の会話構造を Y とする．これらの会話構造を図 8 に示す．

実際の会話を参照したところ，X で発表者が発話した 3 件の場合は，いずれも発表者の発話後に被発表者が質問を続けていた．Y で発表者が発話した 10 件のうち，8 件は質問者が質問を終えて発表者の応答に移ったところであり，残り 2 件でのみ被発表者が質問を続けていた．このことから，発表者は視線によって発話の権利を制御しており，それがこの木構造の差に現れているのではないかと考えられる．

5.2 ポスター環境自由会話で得られた会話構造

ポスター環境自由会話で得られたラベルデータをもとに，ラベルの前処理と抽出結果の検証を繰り返しながらインタラクションマイニングに取り組んだ結果，以下のような会話構造を得ることができた．

(2-1) 発話者は非発話者よりうなずきを行いやすい
(2-2) 発話者は非発話者より視線をポスターから外すことが多い
(2-3) 発話者は非発話者より指差しを行いやすい
(2-4) 2 人と 1 人が異なるポスターを参照している場合，1 人でいる方が発話することが多い

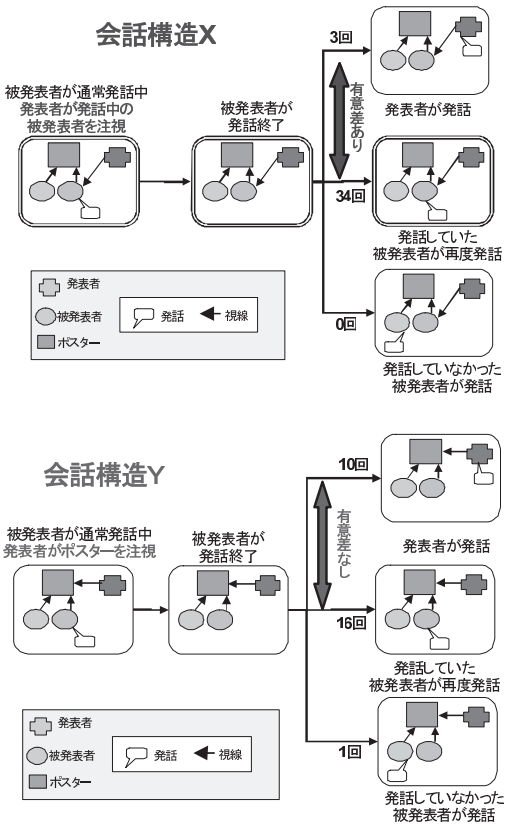

図 8　発表者の視線による被発表者の連続発話数の差異
Fig. 8　Conversation structure about speech by listeners.

(2-5)　指差しを行っている状態では連続発話をする場合が多い

(2-6)　うなずきを行った人物は他の参加者よりも相槌を行う場合が多い

(2-7)　複数の参加者の発話が重複した場合は，先に話していた方が発話をやめる場合が多い

本論文ではこれらのうち (2-4) を取り上げる．(2-4) で得られた会話構造を図 9 に示す．

この構造がより多く見られた session1 に対して，この会話構造が発生している場面を動画と音声で検討したところ，別のポスターを見ている 1 人が発話した場面 15 件中 7 件は「別のポスターで見つけたものをほかの 2 人に知らせる」などの，注視中のポスターに注意を向ける場面であった．4 件はほかの 2 人の方を見る際の遷移中に発生したごく短い区間，4 件はよそ見をしながら話をしている場合などであった．一方，2 人でポスターを見ている側が発話した場面のほとんどは，3 人が同じポスターを注視している場面から視線遷移中のごく短い区間であった．このことから，この構造はほかの会話参加者に自身の注目するものに対して注意を促すものであると考えられる．

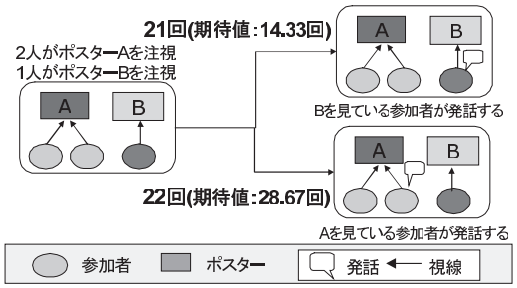

図 9　視線が二つのポスターに分かれているときの発話に関する会話構造
Fig. 9　Conversation structure about speech when subjects' gaze are distributed to two posters.

6.　異なる会話環境の会話構造比較

ポスター発表会話及びポスター環境自由会話についてインタラクションマイニングを行った上で，得られた会話構造を比較した．なお，両者を「ポスターがある環境での 3 人会話」という形に抽象化するため，ポスター発表会話では発表者・被発表者の区別を行わず同一として扱った．

まず，共通して出現した会話構造としては以下のようなものが得られた．

（1）　発話者は非発話者より指差しが多い

（2）　発話者は非発話者より視線を外しやすい

（3）　視線を外していた人物がポスターに視線を戻した後，再びポスターから視線を外す回数がほかの人物より多い

（4）　指差しを行っている人物は発話が多い

（5）　うなずきを行っている人物はその後に相槌を行うことが多い

これらは，3 人がポスターを介して行う会話では共通して出現する会話構造であると考えられる．一方で，以下のものについては会話構造に差異が生まれた．

（1）　ポスター発表では，会話参加者 A が発話終了して沈黙した後，最初に話し始めるのは A である場合が有意に多いのに対し，ポスター環境自由会話では有意差は見られない．

（2）　ポスター発表では，指差しを行っている人物は視線を外す回数が有意に多いのに対し，ポスター環

境自由会話では有意差は見られない．

これらの差異が生じた理由について検証する．まず，（1）は，二つの会話環境における発話交代の頻度差を表していると考えられる．（2）については，ポスター環境自由会話では他の参加者の発話から自分の指差し行為が理解されているかを確かめられるのに対し，ポスター発表では聞き手の視線やうなずきといった非言語行動からそれを知る必要があるためではないかと考えられる．

7. むすび

本研究では，多数の非言語行動の間に存在する会話構造に着目し，これをデータマイニング的な手法で抽出するインタラクションマイニングを提案した．ポスター発表会話とポスター環境自由会話という二つの会話環境におけるデータに対し本手法を適用し，二つの会話環境それぞれに存在する会話構造を抽出することを試みた．また，収録した異なる二つの会話環境における会話構造を比較した．その結果，ポスター発表会話・ポスター環境自由会話のそれぞれで，様々な会話構造を数多く抽出することができた．また，二つの異なる会話環境において共通する会話構造と異なる会話構造を挙げることができた．

インタラクションマイニングの今後の課題としては，まず手法の改良が挙げられる．現在の手法では生起回数の少ない相槌などに関する会話構造が抽出しにくいといった点や，ごくあたり前の会話構造に影響されて複雑な会話構造が抽出しにくいといった面がある．これを改善するため，森田らの研究 [7] で行われているような回数の正規化手法や，既知のルールの除去手法などを検討していきたいと考えている．更に，分析の基盤となる会話ラベルを効率的に，かつ高精度で取得するため，非言語行動のセンシング精度向上に今後取り組んでいきたいと考えている．

最後に，インタラクションマイニングの適用範囲と限界について考察する．インタラクションマイニングは，ポスター会話だけなく，着座式のミーティング会話や，会話をしながらの共同作業のような状況にも適用できると考えている．その一方で，会話参加人数が多くて会話場が複数に分裂してしまうような状況には適用が困難であると想像される．なぜなら，同時刻に生じたからといって，複数の会話場で別々に発生したインタラクションに関するデータでインタラクションステートを作成しても，有意な構造抽出は難しいと考えられるからである．そういった複雑な会話状況に本手法を適用するには，まず前段階として動的に変化する会話場を特定することが必要である．会話場検出のための筆者らの先行研究 [14], [15] を適用して，そういった複雑な会話状況への適用にも挑戦したい．

謝辞　本論文で紹介したインタラクションマイニングの初期の実装は福間良平氏（現在，奈良先端科学大学院大学所属）によってなされた．会話データ収録とラベリングは，筆者らと同じ研究室に所属する勝木弘，矢野正治，齊賀弘泰の各氏に加え，京都大学学術情報メディアセンターの河原達也教授，高梨克也助教と議論・協力しながら進めた．以上の各氏に深く感謝する．本研究は，文部科学省科学研究費補助金「情報爆発時代に向けた新しい IT 基盤技術の研究」の一環で実施された．

文　献

[1] 坊農真弓，鈴木紀子，片桐恭弘，“多人数会話における参与構造分析 — インタラクション行動から興味対象を抽出する，” 認知科学，vol.11, no.3, pp.214–227, 2004.

[2] 榎本美香，伝　康晴，“3 人会話における参与役割の交替に関わる非言語的行動の分析，” 人工知能学会研究会資料，no.SIG-SLUD-A301, pp.25–30, 2003.

[3] D. McNeill, “Gesture, gaze, and ground,” Second International Workshop on Machine Learning for Multimodal Interaction (MLMI 2005), Lect. Notes Comput. Sci., vol.3869, pp.1–14, Springer, 2006.

[4] メイナード K 泉子，“日米会話における非言語行動の対照分析 — 頭の動きをめぐって，” 会話分析，第 9 章，pp.167–179，くろしお出版，1993.

[5] J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, I. McCowan, W. Post, D. Reidsma, and P. Wellner, “The AMI meeting corpus: A pre-announcement,” Second International Workshop on Machine Learning for Multimodal Interaction (MLMI 2005), Lect. Notes Comput. Sci., vol.3869, pp.28–39, Springer, 2006.

[6] L. Chen, R. Rose, Y. Qiao, I. Kimbara, F. Parrill, H. Welji, T.X. Han, J. Tu, Z. Huang, M.P. Harper, F.K.H. Quek, Y. Xiong, D. McNeill, R. Tuttle, and T.S. Huang, “VACE multimodal meeting corpus,” Second International Workshop on Machine Learning for Multimodal Interaction (MLMI 2005), Lect. Notes Comput. Sci., vol.3869, pp.40–51, Springer, 2006.

[7] 森田友幸，平野　靖，角　康之，梶田将司，間瀬健二，萩田紀博，“マルチモーダルインタラクション記録からのパターン発見手法，” 情処学論，vol.47, no.1, pp.121–130, 2006.

[8] 大塚和弘，竹前嘉修，大和淳司，村瀬　洋，“複数人物の対

面会話を対象としたマルコフ切替モデルに基づく会話構造の確率的推論," 情処学論, vol.47, no.7, pp.2317–2334, 2006.
[9] 独立行政法人国立国語研究所, "日本語話し言葉コーパス," 2008. http://www.kokken.go.jp/katsudo/seika/corpus/
[10] 吉田奈央, 高梨克也, 伝 康晴, "対話におけるあいづち表現の認定とその問題点について," 言語処理学会第 15 回年次大会発表論文集, pp.430–433, 2009.
[11] A. Kendon, Gesture: Visible Action as Utterance, Cambridge University Press, 2004.
[12] 福間良平, 角 康之, 西田豊明, "人のインタラクションに関するマルチモーダルデータからの時間構造発見," 情処学研報（ユビキタスコンピューティングシステム）, vol.2009-UBI-23, 2009.
[13] 齊賀弘泰, 角 康之, 西田豊明, "多人数会話におけるうなずきの会話制御としての機能分析," 情処学研報（ユビキタスコンピューティングシステム）, vol.2010-UBI-26, 2010.
[14] 高橋昌史, 角 康之, 伊藤禎宣, 間瀬健二, 小暮 潔, 西田豊明, "時系列イベント発見のためのグラフクラスタリング手法の提案," 情処学論, vol.49, no.6, pp.1942–1963, 2008.
[15] T. Nakakura, Y. Sumi, and T. Nishida, "Neary: Conversation field detection based on similarity of auditory situation," Tenth Workshop on Mobile Computing Systems and Applications (HotMobile 2009), 2009.

（平成 22 年 3 月 28 日受付，7 月 21 日再受付）

中田 篤志

2008 京大・工・情報卒．2010 同大大学院情報学研究科修士課程了．在学中，実世界インタラクションの計測と理解の研究に従事．現在は，（株）NTT データに勤務．

角 康之 （正員）

1990 早大・理工・電子通信卒．1995 東京大学大学院工学系研究科情報工学専攻了．同年，（株）国際電気通信基礎技術研究所（ATR）入所．2003 より，京都大学大学院情報学研究科助教授（現在，准教授）．博士（工学）．研究の興味は実世界インタラクションの計測・理解・支援と体験メディア．

西田 豊明 （正員）

1977 京大・工・卒．1979 同大大学院修士課程了．1993 奈良先端科学技術大学院大学教授，1999 東京大学大学院工学系研究科教授，2001 東京大学大学院情報理工学系研究科教授を経て，2004 年 4 月京都大学大学院情報学研究科教授，現在に至る．会話情報学，原初知識モデル，社会知のデザインの研究に従事．日本学術会議連携会員 (2006〜)，人工知能学会会長 (2010〜)，国立情報学研究所運営会議委員 (2008〜)．