社団法人　情報処理学会　研究報告
IPSJ SIG Technical Reports

2004－NL－164（22）
2004／11／5

事象の認識による発話生成に向けて

小南　光† 齋藤　豪†† 奥村　学†††

† 東京工業大学大学院総合理工学研究科知能システム科学専攻
226-8503 神奈川県横浜市緑区長津田町 4259
†† 東京工業大学大学院情報理工学研究科計算工学専攻
152-8550 東京都目黒区大岡山 2-12-1
††† 東京工業大学精密工学研究所
226-8503 神奈川県横浜市緑区長津田町 4259
E-mail: †konan@lr.pi.titech.ac.jp, ††suguru@img.cs.titech.ac.jp, †††oku@pi.titech.ac.jp

あらまし　人は刻々と変化する状況に対して、その変化を認識し、他の人にそれを伝達する場合、その中から必要なものを選択して適切な表現に変換して行っている。このような技術は、今後、人とコンピュータのインタフェースとして必要な技術である。そこで我々は、この状況の変化（事象）を端的に表現することをコンピュータで実現する手法を提案する。本研究では、特に、状況の変化を伝えるために認識した事象の中からどれを選択するかを、注目している時間（注目時間幅）と、認識した事象からその状況を表現する事象を選択するタイミング（選択タイミング）の観点から論じる。人間が注目時間に対して、認識した状況の変化をどのように選択しているかを模倣したモデルを作成し、そのモデルの妥当性と有効性を検証する。
キーワード　事象認識，注目時間幅，選択タイミング

Towards Utterance Generation by Event Recognition

Hikaru KOMINAMI†, Suguru SAITO††, and Manabu OKUMURA†††

† Department of Computational Intelligence and System Science,Interdiciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology
4259 Nagatsuta-cho Midori-ku Yokohama 226-8503 Japan
†† Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology
2-12-1 Ookayama Meguro-ku Tokyo 152-8550 Japan
††† Precision and Intelligence Laboratory, Tokyo Institute of Technology
4259 Nagatsuta-cho Midori-ku Yokohama 226-8503 Japan
E-mail: †konan@lr.pi.titech.ac.jp, ††suguru@img.cs.titech.ac.jp, †††oku@pi.titech.ac.jp

Abstract　In telling what happened, we choose the core phenomena among the recognized various information in the situation and generate appropriate sentences for them. This ability would be required for computer systems. In this paper, we propose the technique to generate expressions for changed situations. Especially, we discuss the width of attention time and the timing of selection of suitable phenomenon in order to report changes of situation. Therefore, we introduce the model that imitates people's way of recognizing changes of situation by the width of attention time. Then we verify the validity and usefulness of the model.
Key words　event recognition, the width of attention time, the timing of selection of suitable phenomenon

1. はじめに

世界の状況は、同時かつ刻々と変化しており、人間はこれらの変化を取捨選択や統合をして認識している。人間が認識する状況の変化には以下のような種類がある。

・物体の移動、変形、消失

・観察主体の移動による相対的変化
・観察主体の記憶との相違から生じた変化
・主体物の意思や目的に基づいた行動

これらの状況の変化を端的に表現することは、この情報を伝達する上で重要な役割を持つ。例えば、実況や視覚障害者のための音声ガイドなどである。実況は、主にサッカーや野球などのスポーツで行われており、選手の動きやボールの行方など、実況者が見たことを取捨選択、統合して視聴者に伝えている。また、視覚障害者のための音声ガイドとは、テレビではドラマにおいて、出演者がテレビ画面内で行っていることの中で、場面の認識を行う上で重要となることを台詞に重ならないように挿入しているもので、実際に画面に映っているものよりも、情報量的には少なく、場面のエッセンスだけを抽出している。

さらに、コンピュータやネットワークが普及し、コンピュータと人間が協調して作業する必要がある現在では、コンピュータが状況の変化を端的な表現で伝達することが求められる。これらのシステムの例として、監視カメラを利用しているセキュリティシステムやロボットの遠隔操作などが挙げられる。監視カメラを利用したセキュリティシステムでは、監視対象となるカメラ映像が膨大な量となるため、この中から異常な状態や不審者などが写っている映像を検索、抽出する手法が求められている。また、レスキューロボットなどの遠隔操作では、そのロボットの周りの状況を的確に判断する必要がある。

このように状況の変化や場面を的確に表現する事は重要な課題である。この課題に対する研究として、実況解説の分野では、サッカーを対象にした [1] や、麻雀を対象にした [2] などがある。これらの研究では、実況の対象となるゲームがあり、そのゲームにおける特徴的な事象を抽出、実況生成を行っている。また、実画像を用いて画像中の人物が何をしているかを、人物とその部屋の中のオブジェクトとの相対的な位置関係から推定、自然言語表現の生成を行っている研究として [3] が挙げられる。この研究は、1 人の行動を時系列に沿って述べていくもので、その人物にのみ焦点が当てられている。

そこで本研究では、状況の変化を表す端的な表現をコンピュータで自動生成する事を目的とし、注目するオブジェクトは状況の変化の流れと発話の流れから変化するものとする。本研究では、「事象があり、これを観察、認識し、発話する」という人の認識過程を模倣したモデルを構築する。つまり、「事象の認識、認識した事象の中から重要な事象を選択する、選択した事象を言語表現に直す」という過程をもつモデルを構築する。本論文では、事象の認識と認識した事象から重要な事象を選択するまでの過程について述べる。本研究で提案する手法では、状況の変化における多数の事象を認識し、その認識した事象の中からその状況の変化を表現する上で重要な事象を選択、発話するようにすることで、表現する対象を 1 つのオブジェクトに限らずに、状況の変化をよりわかりやすく表現することを可能にしている。また、実世界を観測して得る事象の情報はノイズが多いため分析が難しく、さらにその観測自体も困難である。そのため、本研究では事象を認識し、表現を生成する過程の表現生成の部分に焦点を当てるために、観測が比較的簡単な仮想世界を実験環境として用いることとする。

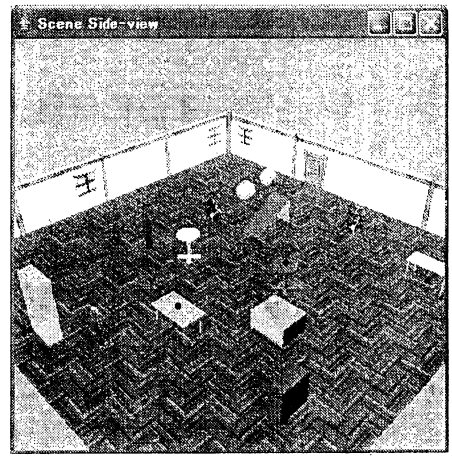

図 1　仮想空間 k3 の俯瞰図

2. 提案手法

実験環境として [4] で開発され、現在も開発が継続している k3 を用いることとする (図 1)。この仮想環境には、テーブルやボール、エージェントなどが配置されており、テーブルやボールといったオブジェクトの物理法則に則った様々な運動を観察することができる。そこで、仮想空間を観測し、オブジェクトに関する様々な値を記録することで、その変化から生じた事象を認識することができる。そのようにして認識した事象集合の中から、その状況に適した表現を生成する手法を提案する。

本研究では、副音声のような音声ガイドのように、仮想空間 k3 における状況の変化をわかりやすく表現することを目指している。図 1 の k3 では、同時に「青いボールが落ちる」「赤いブロックが倒れる」などの事象が観察される。まず、これらの事象を正しく認識する必要がある。次に、このようにして認識された事象は、同時点でいくつも存在することになるので、その中で状況の変化としてもっとも大きな影響を与えている事象、例えば「青いボールが落ちる」を選択する。そして、この選択された事象を最後に適切な表現、例えば「青いボールが青いテーブルから落ちる」のように変換する。

次節以降では、この手法のそれぞれの部分、「事象の認識」「認識した事象の統合」「認識した事象の選択」「表現の生成」について詳しく説明していく。

2.1 事象の認識

本研究で提案するモデルの一番最初の部分、事象の認識部分について述べる。仮想空間を観測し、オブジェクトに関する様々な値を記録することで、その変化から生じた事象を認識する。具体的に観測するものは、オブジェクトの座標、速度、加速度、衝突判定、回転行列の変化である。これらの値の変化からどのような運動と運動の結果残存を認識できるかを以下に述べる。

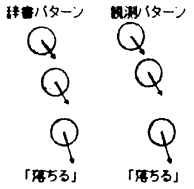

図 2　DP マッチングの例

2.1.1　移動運動

日本語は動詞を経路で考えるという視点 [5] に基づき、オブジェクトの速度と加速度を観測し、その変化量からオブジェクトの運動を認識する。認識する具体的な手法を以下に述べる。

まず、認識したい運動の動詞とその動詞の速度と加速度の変化パターンを登録した辞書を作成する。このとき、運動によって認識するのに必要となる時間が異なるため、運動毎に時間が短い運動か、長い運動かを決めておき、その期間の長さ分だけ速度と加速度の変化量を登録しておく。次に、仮想空間上の各オブジェクト毎に速度と加速度を記録する。そして、その記録された各オブジェクト毎の速度と加速度を、辞書に登録されている各動詞の速度と加速度の変化のパターンとマッチングを取る。このとき、マッチング手法は DP マッチングを用いる。DP マッチングに用いるスコアは、速度どうし、加速度どうしでコサイン類似度をとり、足し合わせたものとする。このようにして求めたスコアが閾値以上であれば、その辞書に登録された動詞の運動として認識する。例えば図 2 のように、辞書に登録されている「落ちる」の速度・加速度の変化パターン (図中左側) と、観測した速度・加速度の変化パターン (図中右側) を DP マッチングにかけて、スコアを計算し、その結果が閾値以上なので、観測したパターンを「落ちる」として認識する。そのため、1 つのオブジェクトのある時間幅に対して認識される運動は 1 つと決まっているわけではない。つまり、時間幅が長い動詞の一部となっている時間幅の短い動詞が、時間幅の長い動詞と一緒に認識されるということがある。

現在、辞書に登録して認識できる移動の運動は、「上がる」「落ちる」「跳ねる」「移動する」「止まる」の 5 つである。

2.1.2　衝突運動

k3 に組み込まれている物理シミュレータ ODE [6] における衝突判定結果を用いて、オブジェクトが他の何かと衝突しているかを判定し、この結果を基に運動を認識する。衝突判定の結果が true ならば、その 2 つのオブジェクトの「ぶつかる」を認識する。この結果を基にして、1 つのオブジェクトに関して、移動運動を認識するときに用いた速度の観測結果を使用して、「ぶつかる」前の速度と「ぶつかった」後の速度で内積を取り、負ならば「跳ね返る」を運動として認識する。また、2 つのオブジェクトがぶつかっていて、片方のオブジェクト A の座標がもう片方のオブジェクト B の平面方向の範囲内にあり、オブジェクト A の垂直方向の座標が、オブジェクト B の高さよりも上にあり、かつ「上に乗る」が認識されていない場合に「上に乗る」を認識する。

2.1.3　回転運動

回転行列をある時間幅で観測し、その変化を調べる。変化があった場合に、その変化量から回転軸を算出することができる。回転行列が変化していて、回転軸が平面方向である場合は「転がる」を認識し、回転軸が平面と垂直な方向である場合は「回る」を認識する。

また、移動運動と衝突運動、回転運動の組み合わせとして、「倒れる」がある。これは、「転がる」が認識されている間「ぶつかる」が認識されており、かつ重心位置が運動の観測の終了時点で、観測の開始点の高さの半分以下になっている場合に「倒れる」を認識する。

2.1.4　運動の結果残存

運動の結果の状態が残り続ける事象が存在する。この結果を、人はその運動の結果残存として認識することができる。この人の認識過程を模倣し、本研究では、結果が残る事象が認識された後、その事象が認識されなくなったとき、その事象の運動の結果残存の事象を認識する。今回実装した動詞が指す運動の中では、「倒れる」「上に乗る」「止まる」が結果が残存する運動にあたる。

以上のようにして、11 種の動詞が表す運動と、運動の結果残存を現在認識することができる。

2.2　認識した事象の統合

各事象の運動の認識には、その運動固有の注目する時間幅が存在する。また、注目時間幅が短い運動を包含する、長い注目時間幅をもつ運動が存在する。この長い注目時間幅を持つ運動が認識されることで、それに包含される短い注目時間幅の運動の認識が、長い注目時間幅を持つ運動の認識に統合される。つまり、同じ運動を観察していても、ある短い時間幅で観察すると、短い注目する時間幅の運動として認識がなされ、長い時間幅で観察すると、長い注目する時間幅をもつ運動として認識される。短い注目時間幅の運動も認識されないわけではないが、表現は抑えられていると考えられる。例を挙げて説明すると、図 3 において、短い注目時間幅を持つ運動の事象 (黒いボール、落ちる) と (黒いボール、上がる)、長い注目時間幅を持つ運動の事象 (黒いボール、跳ねる) が認識されている。このとき「跳ねるかつ上がる」とは言わない。従って、事象 (黒いボール、跳ねる) によって、事象 (黒いボール、落ちる) と (黒いボール、上がる) が統合される。

同様の関係は「倒れる」と「転がる」にも言えて、短い時間幅では「倒れる」と認識がされるが長い時間幅では「転がる」と認識がなされ、「倒れる」は表現されない。

今回は 2 つの運動の組において、認識した事象の統合を行う。

2.3　認識した事象の選択

次に認識した事象の集合の中から状況の変化を表現する上で重要な事象を選択する方法について述べる。認識される事象は、時間を限定したとしてもオブジェクトごとに認識するため、また同じオブジェクトに対して同時期に複数の事象の認識を許容しているので、数多く存在することとなる。この中から、ある状況を表現する際に重要と考えられる事象を選び出す作業が必

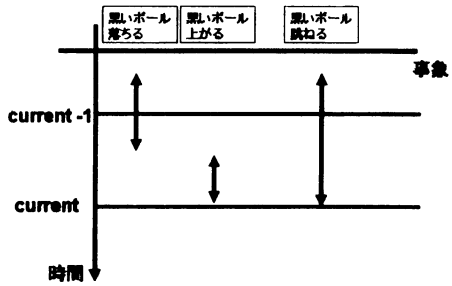

図 3　認識した事象が統合される例

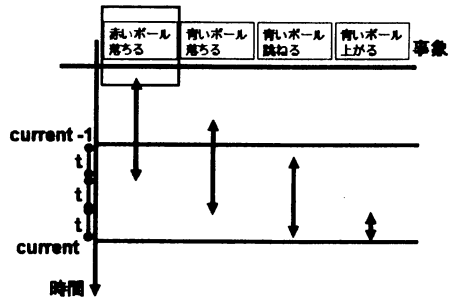

図 4　選択するタイミングを決定する例

要となる。そこで、状況の変化に寄与した事象集合から重要な事象を選択するにあたって、人間が考慮していることをもとに、状況の変化をわかりやすく伝えるためのモデルを構築する。人間が事象を認識するときには、観察している時間や視野に入っているかどうか、これまでの発話の文脈などを考慮して行っていると考えられ、本研究のモデルを構築する場合にも、これらの情報を組み合わせる必要がある。

本論文では、人が考慮している情報の中で、観察する時間のみに注目し、モデルを構築する。このモデルでは、観察する時間幅と発話を選択するタイミング、そして前回発話選択をしたタイミングと現在の発話を選択するタイミングの間に含まれている事象の状況の変化量から、現在の状況の変化を伝える上で重要な事象を選択する。具体的には、観察をしている中で、常に状況の変化量をオブジェクト毎に合計を計算し、これが閾値以上になったときに事象の選択を行う。この閾値は、人間が事象を観察しているときに、ある程度大きな変化でないと認識・発話されないことに基づいて考案した。

まず、認識した事象は、以下の素性をもつものとし、認識した事象のリストに格納する。

- **主格**：認識した状況の変化で、その運動を表す動詞の主語となるオブジェクト
- **主動詞**：認識した状況の変化で、その運動を表す動詞
- **目的格**：認識した状況の変化で、その運動を表す動詞の目的格にくるオブジェクト
- **事象の認識の開始時間**：認識した状況の変化で、その運動を認識できた時刻の開始点
- **事象の認識の終了時間**：認識した状況の変化で、その運動を認識できた時刻の終了点
- **状況の変化量**：認識したこの事象が状況に対してどれほどの影響力を持つかを示す値。この値は認識できる事象の主動詞によって決定される値で、変化しないものとする。

以上の素性のうち、主格、目的格、主動詞、状況の変化量は、事象が認識された時点で決定される素性である。そして、事象の認識の開始時刻は、認識された事象リストの中に、同じ主格、目的格、主動詞を持つ事象が存在せずに新たに認識されたときとする。

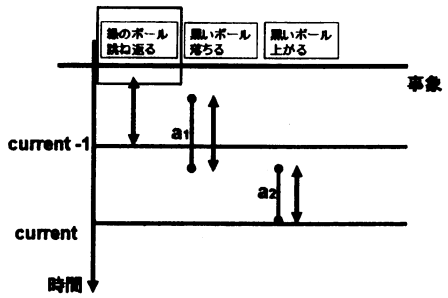

図 5　事象毎の注目する時間幅を表した例

2.3.1　表現を生成する事象を選択するタイミングの決定

図 4 は、認識した事象の中から状況の変化を表現する上で重要な選択をするタイミングを決定する例を表している。図 4 において、各長方形は事象の主語と動詞をそれぞれ表している。各事象の下の矢印は、その事象が認識された時間を表している。また、現在の選択タイミングを current で、前回の選択タイミングを current-1 で表している。まず、前回の選択したタイミング current-1 から微小時間 t づつ観察時間を延ばしていく。そして、その観察時間内で事象の素性である主格のオブジェクト毎に状況の変化量の合計を計算し、これが閾値を超えたときに、その時点を選択するタイミング current とする。

この各オブジェクト毎に状況の変化量の合計を計算するときには、次の 3 つのルールを適用する。

ルール 1：前回の選択タイミングと注目時間の関係

注目する時間幅は、認識する事象毎に異なるため、各事象固有の時間幅を用いることに注意する。こうすることで、前回の選択タイミングから今回の選択タイミングの間に注目する時間幅の一部しか含まれていない事象も、重要な事象として選択・表現することが可能となる。例えば、図 5 の前回の選択タイミングで (緑のボール、跳ね返る) を選択したため、表現できなかった (黒いボール、落ちる) のようにが次の選択タイミングで候補となり、表現される可能性が生じることになる。

ルール２：結果残存の事象の状況の変化量

結果残存の事象の認識は、その原因となった運動の事象の認識が終了した後に行われるもので、状況に変化を及ぼしていないため状況の変化量を定められないが、近似的にその原因となった運動の事象がもつ状況の変化量をもつこととする。

ルール３：継続時間の長い事象の状況の変化量

人は、状況に大きな変化が生じたときに、その状況の変化を伝えようと表現を生成する。このときに、一度表現した事象がその後も継続している場合や、「時計の針が動く」のように長時間にわたって継続されている事象を、人は認識しているが表現する事象の候補から除外して考えている。そこで、一度表現した後も継続している事象や長時間にわたって継続されている事象は、その事象の主格となるオブジェクト毎の状況の変化量の合計に対する閾値を上げる。

また、統合した事象の運動と、その運動に包含される運動の事象の間の関係にも注意する。前回の表現の選択タイミングで、2.2 節で挙げた短い注目時間幅の運動の事象を統合した事象を選択・表現している場合に、現在の表現の選択タイミングで、その事象に包含される短い注目時間幅の運動の事象を表現することはあまり見受けられない。このような場合も上記のようにオブジェクト毎の状況の変化量の合計に対する閾値を上げる。

具体例で説明する。図 4 では、まず t をずらして状況の変化量を計算すると、前回のタイミングで認識されていた事象 (赤いボール, 落ちる) と事象 (青いボール, 落ちる) が認識されるが、各オブジェクト毎に状況の変化量の合計を計算すると、閾値を超えないので選択しない。さらに t をずらし 2t にする。このとき認識されているのは、t の時と同じなので事象の選択はしない。さらに観測時間を延ばして 3t にする。このとき事象 (青いボール, 上がる) と (青いボール, 跳ねる) が追加され、認識した事象が (青いボール, 跳ねる) に統合されるが、状況の変化量の合計が閾値を超えるので、この時点を表現する事象を選択するタイミングとする。またこのとき、このタイミングを決定したオブジェクト「青いボール」に関する事象から表現する事象を選択する。

2.3.2 表現する事象の選択

前節で表現するタイミングを決定するのに貢献したオブジェクトに関する事象の中で、もっとも大きな状況の変化量を持つ事象を表現として選択する。

例を挙げて説明すると、図 4 の例において、前節で表現を選択するタイミングの決定に関わった、オブジェクト「青いボール」に関する事象を選択することが決まり、その (青いボール, 跳ねる) が表現する事象として選択される。

2.4 発話表現の生成

最後に、本研究のモデルの最後の部分である表現の生成について述べる。発話すると決まめられた認識された事象は、事象の素性の中の主格、主動詞、目的格の３つと、認識した事象が結果残存の場合は、主動詞の語尾を「ている」に変換するというルールから単文を生成する。

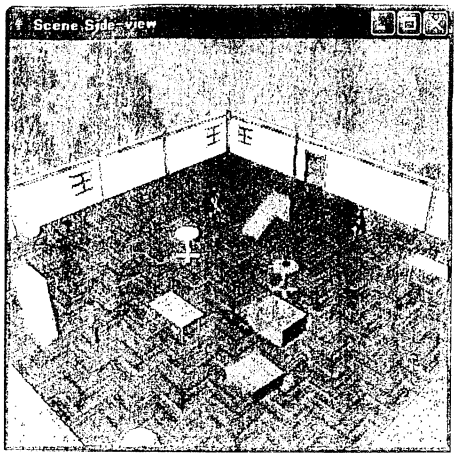

図 6　仮想空間 k3 のオブジェクトが移動した様子

(緑のボール、落ちる)
(緑のボール、ぶつかる、滑り台)
(赤いブロック、倒れる)
(緑のボール、転がる)
(ピンクのボール、転がる)
(青いボール、落ちる)
(青いボール、転がる)
(青いボール、ぶつかる、赤いブロック)
(青いブロック、倒れる)
⋮

図 7　図 1 の状態から図 6 の変化で生じた事象

(黒いボール、上に乗る (結果残存)、赤い丸いテーブル)
(緑のボール、転がる)
(青いボール、跳ねる)
(青いブロック、倒れる)
(青いブロック、跳ねる)
(緑のボール、跳ね返る)
(ピンクのボール、跳ね返る)
(青いボール、転がる)
(緑のボール、転がる)
(青いボール、跳ね返る)
⋮

図 8　システムが認識・選択した事象

3. 事象の認識・選択に関する実験と考察

図 1 の各オブジェクトに力を加えることで、オブジェクトを動かし、図 6 の状態に変化させ、これを観察・認識する。図 1 から図 6 への変化としては、図 1 の中央上部のボールが転がっての移動や、図 1 右下部のブロックが倒れるなど、図 7 に挙げた事象が生じている。

このときシステムが認識・選択した事象を図 8 に示す。この結果と生じた事象を比較する。

この実験では、仮想空間上に15個のオブジェクトを配置し、そのうちの5つに力を加えて動かして行った。ここで、状況の変化を表現する上で重要な事象と考えられるのが、およそ20個あり、そのうち本研究のシステムが認識し、選択した事象が13個であった。また、表現する事象を選択するタイミングにおいては、最も多い場合で50個ほどの事象が認識されており、タイミングを決定した事象の主格となるオブジェクトで5、6個の事象に絞込み、そこから状況の重要な変化を表す事象1つを選択することができた。これらの結果から、本論文で提案した認識した事象から重要な事象を選択するモデルが、ある程度有効であると考えられる。次に今回のモデルにおいて、選択する必要があると考えられるが選択されなかった事象について考察する。まず、本手法で提案していた結果残存の事象選択が何も運動がない最初の「黒いボールが赤い丸いテーブルの上に乗っている」だけしかない。図7にある(赤いブロック、倒れる)「赤いブロックが倒れる」が図8では表現されていないので(赤いブロック、倒れる(結果残存))「赤いブロックが倒れている」と出力されるべきだが、実際には他のオブジェクトの運動によって選択されていない。事象の状況に与える変化量の決め方を、他の事象の動詞も含めてもう少し吟味する必要がある。

また、実際に生じた時系列を考えてみると、図7に示してある事象は、一連の状況の変化がわかるが、図8では、個々の事象はある程度重要かもしれないが、その事象が生じた原因となる重要な事象が抜けていることが多い。「緑のボールが転がる」は「緑のボールが滑り台の上に落ちた」からである。また、「ピンクのボールが跳ね返る」では、ピンクのボールが運動していることが選択されておらず、突然登場するのは、不自然である。これは、本研究で提案したモデルが事象の認識から発話表現の生成にいたる過程をone pathで構築していることに原因がある。このモデルでは、上記の例のように因果関係などから、後になって重要とわかる事象を表現する手段が無いからである。今後この方法を考える必要がある。また、このような現象が生じるのは、本手法ではほぼ同時刻に生じた事象も選択できるように工夫して提案したが、その工夫が不十分であったことも原因と考えられる。図7に示した事象はほぼ同時刻に生じた事象であるが、図8では、それが上の方の4つほどしか示されていない。このように、ほぼ同時刻に起きた事象をあまり選択することができていない。この点を改善する必要がある。

4. ま と め

状況の変化を伝達する上で、状況の変化の中の事象から適切な事象を選び出すことが最も重要である。事象の選択にあたっては、いろいろな情報を加味する必要があるが、本論文では特に、事象の選択するタイミングと事象を認識するのに必要な注目時間幅を考慮して選択するモデルを提案した。このモデルでは、大きな状況の変化が新たに生じたときに必要に応じて、事象の選択ができるようになっている。このことで、状況に変化は生じてはいるが、改めて特筆するべき事柄がないような場合でも、事象を選択してしまわないようになり、必要な情報だけを伝達することを可能にしている。

本研究では、このモデルを使用した事象の選択手法を提案、実装し、事象の認識・選択実験を行った。実験の結果、選択タイミングと注目する時間幅の関係をもとにした本手法が目指していた、前回の選択タイミングで選んだ事象との関係から今回選択する事象が変化することを確認することができた。しかし、同時に生じる事象の表現に限界があることがわかった。また、今回は注目時間にのみ着目したモデルであったが、人間が認識の過程に用いている他の技術を考慮することでよりよいモデルが作成できると考えられる。

以上のことをふまえて今後の課題について述べる。

まず事象を表現するとき、事象の素性と簡単なルールに基づいた単文しか生成していないので、これを改善する。現在考えているのは、周辺情報も組み込んでよりわかりやすい表現を作成することである。例えば、「青いボールが転がる」ではなく、「青いボールが白いテーブルの前を転がる」のように表現するということである。現在は、周辺の情報までは記録していないので、周辺情報も記録できるように改善する必要がある。

また、よりよいモデルの作成のために、次のことを考慮する必要がある。まず、因果関係などから後になって重要とわかる事象の表現を発話表現として選択できるようにする。

そして、選択タイミングや注目する時間幅以外の条件として、視野に関する情報を加える必要がある。視野に関する情報としては、人間は視野の中心に近いほどよく見えているので、視野の中心に近いオブジェクトに関する事象ほど選択されやすい、つまり、状況の変化量が大きいと考えられる。また、視野内にどれだけの大きさでそのオブジェクトが映っているかも、事象を認識する上では大切な情報である。視野の中心付近にあっても小さくしか見えていないオブジェクトに関しては、事象が認識しにくいからである。よって、視野内に大きく映っているオブジェクトに関する事象ほど変化量が大きいとする。さらに、人間は視野の中心である視点を動かして観察しているが、この視点をこれまでの文脈から現在注目度が高いオブジェクトに移して観察している。つまり、注目度が高いオブジェクトに関する事象を認識しやすいようにしているのである。このことを考慮して、文脈から注目度の高いオブジェクトを推定し、この情報も使用することを考えている。

文 献

[1] Tanaka-Ishii,K.,Hashida,K.,and Noda, I.: Reactive content selection in the generation of real-time soccer commentary, in Proceedings of the 17th International Conference on Computational Linguistics, pp.1282-1288,1998

[2] 藤沢瑞樹, 齋藤豪, 奥村学：情報量の異なる複数視点を考慮した実況解説の自動生成, 人工知能学会論文誌, Vol.19, No.6, 2004

[3] 小島篤博：映像中の人物行動の認識とその自然言語記述に関する研究, 大阪府立大学博士論文, Jul., 2003

[4] 新山祐介, 秋山英久, 鈴木泰山, 徳永健伸, 田中穂積：自然言語を理解するアニメーテッドエージェントのための3次元仮想空間における位置の表現と処理, 第13回人工知能学会全国大会論文集, pp.217-220,1999

[5] 松本 曜「認知意味論」シリーズ認知言語学入門3巻 (大修館書店,2003)

[6] http://ode.org/