

知覚的群化に基づく参照表現の生成

船越 孝太郎[†]

渡辺 聖[†]

栗山 直子^{††}

徳永 健伸[†]

参照表現とは、複数の物体のなかから特定の物体を識別するための言語表現である。これまでに提案されている参照表現の生成手法は、物体の属性と物体間の二項関係のみを用いていた。しかしながら、このようなアプローチでは、弁別的な属性や二項関係が存在しない状況において適当な参照表現を生成できない。この問題を克服するため、本論文では、知覚的群化と多項間関係を用いた参照表現の生成手法を提案する。人間が知覚的に認識しやすい物体の群を利用することで、「左の三つの玉のうちの一番右の玉」のような表現を生成することが可能になる。我々は 42 人の被験者に対して心理実験を行ない、知覚的群化を用いた参照表現を収集した。生成アルゴリズムは収集した表現に基づいて構築した。23 人の被験者を用いた評価実験により、提案手法が適切な参照表現を生成できることが確認できた。

キーワード: 参照表現, 空間記述, 知覚的群化, 言語生成

Generating Referring Expressions based-on Perceptual Grouping

KOTARO FUNAKOSHI[†] SATORU WATANABE[†] NAKO KURIYAMA^{††}
and TAKENOBU TOKUNAGA[†]

Past work of generating referring expressions mainly utilized attributes of objects and binary relations between objects to distinguish the referent from other objects. However, such an approach does not work well when there is no distinctive attribute among objects. To overcome this limitation, this paper proposes a novel generation method utilizing the perceptual groups of objects and n -ary relations among them. With the proposed method, an expression like “the rightmost ball in the left cluster of three balls” can be generated. The key is to identify groups of objects that are naturally recognized by humans. We conducted psychological experiments with 42 subjects to collect referring expressions in such situations, and built a generation algorithm based on the results. The evaluation using another 23 subjects showed that the proposed method could effectively generate proper referring expressions.

KeyWords: *Referring Expression, Spatial Description, Perceptual Grouping, Language Generation*

[†] 東京工業大学 大学院情報理工学研究科 計算工学専攻

Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology

^{††} 東京工業大学 大学院社会理工学研究科 人間行動システム専攻

Department of Human System Science, Graduate School of Decision Science and Technology, Tokyo Institute of Technology

1 はじめに

過去 20 年の間、機械と人とのコミュニケーションを実現するために、多くの研究者が参照表現の生成に取り組んできた。参照表現の生成において、先行研究の多くは、聞き手に特定させたい対象（以後、対象物とよぶ）の属性、および対象物と他の物体（以後、非対象物とよぶ）の間の二項間関係を用いている (Appelt 1985; Dale and Haddock 1991; Dale 1992; Dale and Reiter 1995; Heeman and Hirst 1995; Horacek 1997; Krahmer and Theune 2002; van Deemter 2002; Krahmer, van Erk and Verleg 2003)。従って、これらの手法は、対象物と非対象物との間に外見的な差異が存在せず、どの二項関係も対象物を特定するには有効でない状況において、適切な参照表現を生成することができない。ここで「適切な参照表現」とは、聞き手に対象物を特定させうる簡潔で自然な表現を指す。後に示すように、論理的には成立するものの、人間に提示する表現としては不適切な表現も存在する。

例として、図 1 において対象物 b を図中の人物 P に示すことを考える。ただし、 a 、 b などのラベルは説明上付加したもので、人物 P には提示されない。対象物 b は外見からは物体 a や物体 c から区別することができない。そこで次の方策として、対象物 b とテーブルとの間の関係を用いることが考えられる（例えば「テーブルの右の玉」）。しかし、物体 a も物体 c もテーブルの右にあるため、この状況においては「 X の右の Y 」という関係に弁別能力は無い。テーブルの代わりに物体 a や物体 c を参照先として使うことも意味が無い。なぜなら物体 a および物体 c は物体 b が一意に特定できないのと同じ理由によって一意に特定することができないからである。このように、物体の属性と二項間関係のみを用いる従来の手法では参照表現の生成に失敗する。手法によっては「玉の前の玉の前の玉」のような論理的には誤りでない表現を生成できるが、適切な参照表現ではない。このような状況は今まで注目されてこなかったが、物体配置のようなドメイン（例えば (Tanaka, Tokunaga and Shinyama 2004)）では容易かつ頻繁に起こりうる。

(van der Sluis and Krahmer 2000) は、図 1 のような状況において、指差しのようなジェスチャを用いることを提案している。しかしながら、指差しや視線は話し手と聞き手の位置関係によっては常に利用できるとは限らない。一方、図 1 に示す状況では、話し手は「一番前の玉」という表現を用いることで、ジェスチャを用いなくても対象物 b を人物 P に対して示すことができる。このような表現を生成するためには、話し手は知覚的に特徴のある物体群を認識し、群に含まれる物体の間の多項間関係を用いる必要がある¹。このような群の認識を知覚的群化 (perceptual grouping) とよぶ。

本論文では、群の要素間の多項間関係を利用した参照表現の生成手法を提案する。提案手法は、Thórisson の手法 (Thórisson 1994) を用いて知覚的群化を行う。本論文では問題に取り組むにあたり、簡単のため図 2 のように均質な物体が乱雑に配置された状況を対象とする。従って、ここで扱う多項間関係は位置関係だけである。一般には、大きさや色についてもその度合

¹ (Krahmer et al. 2003) は彼らの手法が多項間関係を用いることができると主張しているが詳細は示されていない。我々は彼らの手法は直接的にはここで示したような状況を扱えないと考える。

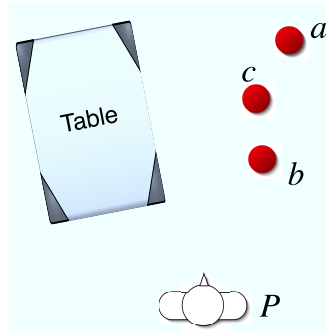


図 1: 従来手法が対応できない状況の一例

いに応じた多項間関係が存在する²。

また，本論文では対話者は適当な参照枠 (Levinson 2003) を共有していると仮定する．参照枠は視点と基準が決まったときに前後・左右の方向を決定する．しかしながら，(Cavazza, Charles and Mead 2002; Tanaka et al. 2004) のような会話エージェントシステムに提案手法を適用しようとするれば，参照枠は対話の各時点において適切に決定されなければならない³。

話し手は対象物を参照する過程でしばしば複数の群に言及する．このような場合，我々は 2 種類の関係を使うことができる．一つは，「机の近くの五つの玉のうちの手前の玉」のような，内部参照関係（「五つの玉」に「手前の玉」が含まれる）である．もう一つは，「五つの玉の右の二つの玉」のような，外部参照関係（「五つの玉」に「二つの玉」は含まれない）である．内部参照関係にある二つの群は必ず包含関係にある．2 節で述べるように，本論文では包含関係だけを用了参照表現に注目する．

本論文の構成は以下の通りである．まず 2 節では，知覚的群化に基づく参照表現を収集するためにに行った実験について説明する．そして提案手法を 3 節に示す．4 節では，知覚的群化の観点から表現の妥当性を予測することが可能かどうかを検討する．最後に，5 節で本論文を結び，今後の課題について述べる．

² 先行研究が「大きさ」などを扱う場合には，相対的な関係としてではなく，例えば「大きい玉」と「小さい玉」のように，個々の物体の離散化された絶対属性として扱っていた．

³ 個々の対象に対する参照枠についての研究は多くなされているが (Clark 1973; Herskovits 1986; Jörding and Wachsmuth 2002; Levinson 2003)，物体の群とその要素に関して参照枠がどのように決定されるのかについてその計算モデルはほとんど研究されておらず，今後検討する必要がある．また，適切な参照表現を生成するためには話者間の視界／視点の違いも考慮しなければならない．問題の一つは，障害物などの存在によって見える範囲が変わることである．もう一つの問題は，人がしばしば 3 次元空間を 2 次元平面に投射して物体を認識するために視点によって位置関係の認識がずれてしまい，例え参照枠を統一してもうまく対象物を特定できない恐れがあることである．

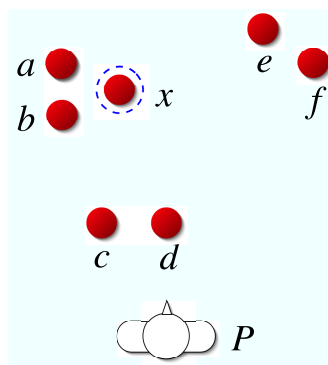


図 2: 視覚刺激の一例（物体のラベルは被験者には見えない）

2 データ収集

知覚的群化に基づく参照表現を収集するために、42 人の日本人大学生に対して心理実験を行った。そして、収集した表現を評価するための実験を、別の 44 人の日本人大学生に対して行った。二つの被験者グループ間に重なりは無い。実験の詳細を以下に記す。

2.1 参照表現の収集（実験 1）

方法: 被験者に同色同形同大の物体が複数配置された 2 次元の俯瞰図を提示し、図中に描かれた第三者に対象物を伝えるように要求した。各図には、3 から 9 個の物体が配置されている。布置は 12 種類を手で作成した。被験者に提示した図の一例を図 2 に示す（提示した 12 種類の布置は、付録 A を参照されたい）。図中の物体に割り振られたラベル $a, \dots, f, x$ は、被験者に示した実際の図では表示されていない⁴。各被験者には、図中の人物 P が破線で囲まれた物体を取ることができるような指令を記述するように指示した。対象物に対する参照表現は、これらの指示表現から取り出した。被験者が適当な表現を作ることができない場合にはその課題を放棄し、次の課題に進むことを許した。対象物は各布置毎に固定したので、全被験者は各布置において同一の対象物に対する言語表現を生成する。

分析: 42 人の被験者に対して 12 の布置を示し、476 表現を収集した。予定した 504 表現 (42×12) のうち、13 表現については被験者に放棄され、15 表現については明らかに誤っている表現であったので取り除いた。収集した表現を分析した結果、被験者は一般にすべての物体を含む群から始めて対象物だけを含む群に絞り込んでいくことを確認した。従って、ある参照表現は被験者の絞り込みの過程を反映した群の系列 (SOG: Sequence of Groups) として形式化できる。

⁴ 以後、 x で対象物を示し、それ以外の小文字英字で非対象物を示す。

以下の例は，図 2 の対象物 x に対して実際に観察された表現である．表現の下に対応する SOG を記す⁵．

「左奥にある三つの玉のうち一番右の玉」

SOG: $[\{a, b, c, d, e, f, x\}, \{a, b, x\}, \{x\}]$

ここで，

$\{a, b, c, d, e, f, x\}$ は図中のすべての物体（全体群）を示し，

$\{a, b, x\}$ は左奥の三つの物体からなる群を示し，

$\{x\}$ は対象物のみからなる群を示す．

絞り込みは全体群から始まるので，SOG はすべての物体の群から始まり，対象物だけを含む群で終わる．収集した言語表現を SOG に翻訳することで，表現を抽象化し分類することができる．各布置に対しては平均 39.7 表現を手に入れたが，それらは平均 8.4 の異なる SOG に分類できた．収集した表現の概要を表 1 に示す．布置が複雑になると，生成される表現が多様化すると同時に放棄や誤りの割合も増えることがわかる⁶．

1 節で述べたように群間の関係は二種類存在する．しかし，内部参照関係（包含関係）だけを用いた表現が収集表現全体の 70% を占めた．

表 1: 収集した表現の概要

布置番号	1	2	3	4	5	6	7	8	9	10	11	12	平均
収集表現の数	41	40	41	41	42	37	42	32	42	41	41	36	39.7
SOG の異なり数	5	6	8	8	6	12	4	15	4	11	5	17	8.4

2.2 収集した表現の評価（実験 2）

方法: 2.1 節で使用した図とその図を用いて収集した表現を被験者に提示し，提示された参照表現が指す物体を選択するように指示した．被験者は，実験 1 の被験者とは異なる日本人大学生の男女である．被験者に提示した図は，対象物を示す破線がないことを除いて，実験 1 で使用したものと同じである．また，被験者には物体の選択に関する確信度と，提示された表現の簡潔さと自然さをそれぞれ 8 段階で主観的に評価させた．

被験者の数が限られているため，評価できる表現の数も限られる．そこで我々は，各布置に対して最大 10 表現を選択して評価を行った．これらの表現は，異なる SOG が可能な限り被覆されるようにした．異なる SOG が 11 以上ある場合には，頻度の高い SOG から順に選んだ．異なる SOG が 10 未満の場合には，頻度の高い SOG から対応する複数の言語表現を選択した．

⁵ 以後，SOG も含めて，群のリストを鍵括弧で閉じて表現する．

⁶ 誤りの割合については，3.3 節の表 3 参照．

ある SOG に対して複数の言語表現がある場合には、頻度の高いもの、あるいは主観的により適切と思われる表現を選んだ。結果、49 人の被験者に 117 表現を評価させた。各被験者は平均して 29.5 表現を評価した。

分析: 無効な回答を除き、全部で 1,429 の評価を得た。一表現あたりの平均評価数は 12.2 である。各表現の質を測るために、評価値という尺度を用意した。評価値は式 (1) によって与えられる。評価値により、どのような表現が好まれるのかを分析した。また、機械生成の表現に対するスコアリング関数 (6) の設計にも用いた (3.2 節参照)。

$$(\text{評価値}) = (\text{正答率}) \times (\text{確信度}) \times \frac{(\text{自然さ}) + (\text{簡潔さ})}{2} \quad (1)$$

分析の結果、内部参照関係だけを用いた 84 表現が高い正答率 (平均 79.3%) と高い評価値 (平均 33.1) を得た。一方、外部参照表現を一度でも用いた 33 表現は、内部参照関係だけを用いた表現に比べ、低い正答率 (平均 69.1%) と低い評価値 (平均 19.7) を得た。外部参照関係を用いない (すなわち内部参照関係だけを用いた) 言語表現は、外部参照関係を用いる表現の 2 倍観察された。図 2 中の対象物 x に対する表現の例を示す。

- 外部参照関係なし
 - 「左奥の三つの玉のうち一番右の玉」
- 外部参照関係あり
 - 「手前の二つの玉の後ろの玉」

また、全体群について言及する表現に対する評価値は全体的に低かった。これらの観察から、我々は内部参照関係だけを用い、全体群については明示的に言及しない表現生成アルゴリズムを構築する。

3 参照表現の生成

本節では知覚的群化に基づく参照表現生成アルゴリズムを提案し、その評価を行う。知覚的群化には Thórisson のアルゴリズム (Thórisson 1994) を採用する。

3.1 Thórisson の知覚的群化アルゴリズム

Thórisson のアルゴリズムは人間の知覚的群化を模倣する。知覚的群化は物体の物理的な近接性、色の類似性、形状の類似性など様々な観点 (特徴) に基づいて行うことができるが、各特徴空間における物体間の「距離」を定義することにより、Thórisson のアルゴリズムは様々な特徴に対して一様に対応する。また、各特徴の距離関数を合成することで、複数の観点に基づいた群化を行うこともできる (例えば、色が類似してかつ近傍に存在する物体だけを群化することができる)。本論文で群化に使用する特徴は近接性だけである。物体間の近接性を表す「距

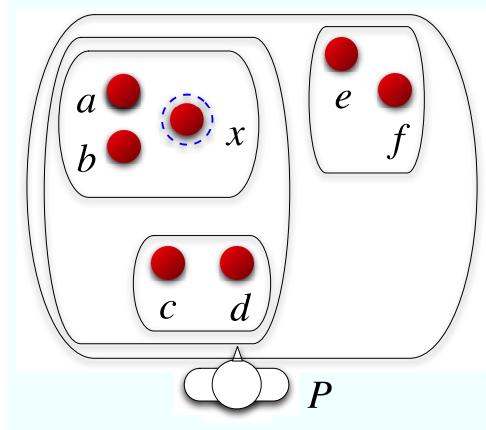


図 3: 群化アルゴリズムによる群化結果

離」は物体の中心間のユークリッド距離として定義した⁷．図 2 の布置に対して群化を行った結果を，図 3 に示す．

Thórisson の群化アルゴリズムは，ある一つの特徴（例えば，本論文の場合では物体の近接性）に関して群化を行う際に，包含関係を持たずに交差する群を生成することはない．3.2 節で提案する表現生成アルゴリズムのステップ 2 の処理は，この特性を前提としている．

Thórisson のアルゴリズムは線状に並んだ物体の群を認識することができない．しかしながらこのような形状は人間には非常に認識しやすいため，人間はしばしば線状の配置に言及する表現を生成する．このような表現も機械が生成・理解するためには，顕現性の高い幾何学形状を認識できる群化アルゴリズムが必要であるが，これについては今後の研究課題とする⁸．

3.2 参照表現生成アルゴリズム

物体の布置と対象物を入力として，提案アルゴリズムは次の 3 ステップで参照表現を生成する．

ステップ 1: 群の列挙: 物体間の近接性に基づいて知覚的群化を行う

ステップ 2: SOG の生成: 群を組み合わせさせて SOG を生成する

ステップ 3: 言語表現の生成: SOG を言語表現に変換する

以降，図 2 を用いて参照表現生成のための上記の 3 ステップを説明する．

⁷ 今回用いる物体は全て均一であるので，物体の大きさは無視した．

⁸ 人間に対して顕現性の高い幾何学形状のほとんどは，辞書やシソーラスの中の形状を表す語彙（例えば，線，円，四角形など）を参照することで数え上げられると予想できる．知覚的群化を模倣するアルゴリズムは，少なくともこれらの語彙によって参照される形状をなす物体配置を認識できる必要がある．

ステップ 1: 群の列挙

物体の布置が与えられると、まず Thórisson のアルゴリズムにより、物体の近接性にもとづいて群のリストを生成する。前節で述べたように、我々は内部参照関係だけを利用するので、生成された群のリストの中から対象物 x を含む群だけを選択する。内部参照関係だけで対象物を特定するためには、SOG を使って対象物を絞り込む過程において、すべての群が対象物 x を含まなければならない。次に、群をその大きさに従って降順にソートする。最後に、もし対象物だけを含む群がソートされたリストに含まれなければ、それをリストの最後に付け加える。以上の処理で得られた群のリスト GL が、ステップ 1 の出力である。

例えば、図 2 に対して、Thórisson のアルゴリズムは五つの群

$$\{a, b, c, d, e, f, x\}, \{a, b, c, d, x\}, \{a, b, x\}, \{c, d\}, \{e, f\}$$

を生成する。対象物を含まない群を除き、対象物だけを含む群を加えて、群のリスト

$$[\{a, b, c, d, e, f, x\}, \{a, b, c, d, x\}, \{a, b, x\}, \{x\}] \quad (2)$$

を得る。

ステップ 2: SOG の生成

このステップでは、ステップ 1 の出力した GL から、2 節で説明した SOG を生成する。SOG は (3) の形式を取る。ここで G_i はある群を表し、 G_0 は全ての物体を含む群である。SOG(3) は、絞り込みが全体群 G_0 から始まり対象物 x のみを含む群で終わることを表している。

$$[G_0, G_1, \dots, G_{m-2}, \{x\}] \quad (3)$$

まず、群のリスト GL から、全ての可能な SOG を生成する。リストのサイズが m のとき、 G_0 と $\{x\}$ は必ず SOG に含まれるので、 2^{m-2} 個の SOG が生成される。従って、群のリスト $[G_0, G_1, G_2, \{x\}]$ からは、四つの SOG $[G_0, \{x\}]$, $[G_0, G_1, \{x\}]$, $[G_0, G_2, \{x\}]$, $[G_0, G_1, G_2, \{x\}]$ が生成される。例えば、リスト (2) から生成される SOG の一つは

$$[\{a, b, c, d, e, f, x\}, \{a, b, x\}, \{x\}] \quad (4)$$

である。3.1 節で述べたように、群化アルゴリズムが生成する群のリスト中の任意の二つの群 G_i および G_j は交差しない ($G_i \cap G_j = \phi$) か、そうでなければ、一方が他方を包含する ($G_i \subset G_j$ あるいは $G_j \subset G_i$)。

ステップ 3: 言語表現の生成

最後に SOG を言語表現に変換する。日本語は主要部後置型言語なので、SOG に含まれる各群に関する言語表現の順序は、最終的な言語表現においても保持される。すなわち、SOG

1. $|G_{i+1}| = 1$ の場合(a) $|G_i| > 2$ の場合

- i. G_{i+1} の BB (バウンディングボックス) が G_i の BB に接触する場合
 $E(R(G_i, G_{i+1})) = \text{「のうちの一番 } X \text{ の」}$: X は, 右, 左, 手前, 奥, 右手前, 左手前, 右奥, 左奥の順に判定して, 最初に該当した関係.
- ii. G_{i+1} の BB が G_i の BB に接触しない場合
 - A. G_i の BB が横長 (図の左右方向に長い) の場合
 対象 x が右から m 番目, 左から n 番目として,

$$E(R(G_i, G_{i+1})) = \begin{cases} \text{「右から } m \text{ 番目の」} & : m \leq n \\ \text{「左から } n \text{ 番目の」} & : m > n \end{cases}$$

- B. G_i の BB が縦長 (図の上下方向に長い) の場合

対象 x が下 (手前) から m 番目, 上 (奥) から n 番目として,

$$E(R(G_i, G_{i+1})) = \begin{cases} \text{「手前から } m \text{ 番目の」} & : m \leq n \\ \text{「奥から } n \text{ 番目の」} & : m > n \end{cases}$$

(b) $|G_i| = 2$ の場合

$E(R(G_i, G_{i+1})) = \text{「のうちの } X \text{」}$: X は, 右, 左, 手前, 奥, 右手前, 左手前, 右奥, 左奥の順に判定して, 最初に該当した関係.

2. $|G_{i+1}| > 1$ の場合

$E(R(G_i, G_{i+1})) = \text{「のうちの } X \text{ の」}$: X は, 右, 左, 手前, 奥, 右手前, 左手前, 右奥, 左奥, 真ん中の順に判定して, 最初に該当した関係.

図 4: $E(R(G_i, G_{i+1}))$ に対する言語表現の付与

$[G_0, G_1, \dots, G_{n-2}, \{x\}]$ は, (5) のように言語化される. ここで, $E(X)$ は X に対する言語表現を表し, $R(X, Y)$ は X と Y の間の関係を示す. 「+」は文字列連結子である.

$$E(G_0) + E(R(G_0, G_1)) + E(G_1) + \dots + E(R(G_{n-2}, \{x\})) + E(\{x\}) \quad (5)$$

ただし, 2.2 節で述べたように, 全体群について明示的に言及する表現の評価値は低かったので, 提案アルゴリズムは $E(G_0)$ については言語化しない. 使用する言語表現は実験 1 で得た実験データから収集した.

言語表現の生成は決定的に行う. 本論文では, $E(G_i)$ ($i > 0$) に対して以下のように言語表現を与える ($n = |G_i|$).

$$E(G_i) = \begin{cases} \text{「} n \text{ つの玉」} & : n > 1 \\ \text{「玉」} & : n = 1 \end{cases}$$

$E(R(G_i, G_{i+1}))$ に対しては, G_i と G_{i+1} それぞれの要素数と互いの位置関係から, 図 4 に示す条件判断に従って言語表現を与える. ただし, $E(R(G_0, G_1))$ については, $E(G_0)$ を言語化しな

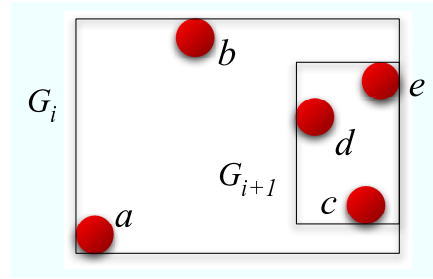


図 5: バウンディングボックス

SOG: $\{\{a, b, c, d, e, f, x\}, \{a, b, x\}, \{x\}\}$

$\rightarrow E(R(\{a, b, c, d, e, f, x\}, \{a, b, x\})) + E(\{a, b, x\}) + E(R(\{a, b, x\}, \{x\})) + E(\{x\})$

$\rightarrow$ 「左奥の」+「三つの玉」+「のうちの一番右の」+「玉」

図 6: 表現生成の例

いため, $E(R(G_0, G_1))$ によって生成される表現から「のうちの」という部分を除く。「 G_i のバウンディングボックス」とは, 図 5 に示すように群の要素を包囲する最小の矩形のことである。位置関係の判定は座標計算によって行う。図 5 における G_i ($\{a, b, c, d, e\}$) と G_{i+1} ($\{c, d, e\}$) の関係は, 図 4 の 2. に相当する。そして二つのグループの位置関係は, G_{i+1} が G_i の右辺に接触しているため「右」に該当すると判定される。

各 SOG に対してアルゴリズムは一つの言語表現を生成する。ただしアルゴリズムは一般に入力配置に対して複数の SOG を生成するので, 最終的に複数の言語表現を生成することがある。SOG (4) は図 6 のように言語化される。

3.3 生成表現の評価 (実験 3)

3.2 節で提案したアルゴリズムを実装し, 実験 1, 2 の被験者とは異なる 23 人の大学生を被験者として評価を行った。評価実験の手順は 2.2 節での実験 2 と同一である。

結果を表 2 に示す。Human-12-all は実験 1 で 12 の布置を用いて人間から収集した表現を実験 2 で評価した結果である。Human-12-90 と Human-12-100 は, それぞれそのうち 90%以上と 100%の正答率を得た言語表現だけに関する評価である。

System-12 は, 実験 1, 2 で用いた 12 布置に対して機械が生成した表現に対する評価の平均値である。12 布置に対して提案アルゴリズムは 18 表現を生成した。これらの 18 表現を各被験者に無作為な順序で提示し, 対象物を特定させるとともに評価を求めた。

System-20 は, 物体を無作為に自動的に配置して生成した 20 の布置に対する平均評価であ

表 2: 生成表現の評価

	正答率 (%)	自然さ	簡潔さ	確信度	評価値
Human-12-all	87.3	4.82	5.27	6.14	29.3
Human-12-90	97.9	5.20	5.62	6.50	35.0
Human-12-100	100	5.36	5.73	6.65	37.2
System-12	91.0	5.60	6.25	6.32	40.1
System-20	88.4	5.09	5.65	6.25	35.2
System-Average	89.2	5.24	5.82	6.27	36.6

表 3: 布置ごとの正答率 (%)

布置番号	1	2	3	4	5	6	7	8	9	10	11	12
Human-12-all	95.8	91.6	95.2	84.0	97.4	74.2	99.2	85.2	100	67.6	97.9	62.7
System-12	100	100	95.7	97.1	100	91.3	100	95.7	97.8	95.7	100	73.9

る．これらの 20 布置については，最低二つの言語表現が生成されるようにした．提案アルゴリズムはこれらの 20 布置に対して 48 表現を生成した．この 48 表現も System-20 の場合と同様に評価した．System-Average は System-12 と System-20 を併せた全体の平均 (micro average) である．

「正答率」の列は，被験者が与えられた表現から対象物を正しく特定できた割合である．前述のように，Human-12-*と System-12 とでは評価を行った被験者集団が異なる．従って数値の直接的な比較はできないが，提案アルゴリズムが適切な表現を生成していることがわかる．

提案アルゴリズムは自然さと簡潔さにおいて，人間が生成した表現より評価が高い．提案アルゴリズムは，実験 1 で収集した表現の中から著者らが自然であると判断した表現を組み合わせで参照表現を生成する．一方，人間が生成した表現には，このような制約がないため，提案アルゴリズムによって生成される表現に比べ多様性があり，中には正答率が高いが簡潔さや自然さにおいて評点の低い表現が含まれている．このような表現が簡潔さ・自然さにおける平均評点を低くしていると考えられる．簡潔さや自然さの評価に負の影響を与えていると思われる性質をこれらの表現に一貫して見いだすことはできないが，聞き手に十分な確信を与えるには簡素すぎる表現 (10 番の布置に対する「真ん中の赤い玉」という表現) や，複雑な布置に対して正確であろうとしたためか冗長になっている表現 (6 番の布置に対する「真ん中あたりに横に並んでいる 4 つの玉の中の左から 2 番目」という表現) などが見られた．

表 3 に，布置ごとの正答率を示す．Human-12-all と System-12 の意味は前述の通りである．12 番の布置においては，人間も機械も正解率が落ちているが，4 番，6 番，8 番，10 番の布置では，人間の作った表現に対する正答率が落ちているのに対し，機械の方は正答率を比較的高く維持できている．

4 SOG に対するスコア付け

前節で提案した表現生成アルゴリズムは、複数の可能な表現を出力する。従って、出力された表現からもっとも望ましいと思われる表現を選択する手段が必要である。生成された表現のスコアは、表現の複雑さ、参照物の顕現性など、いくつかの指標に従って計算することができる。本節では、絞り込みの過程の妥当性を予測できるかどうか、すなわち、SOG のスコアを計算することができるかどうかという問題について検討する。

4.1 群間の規模比によるスコア付け

SOG は (3) に示した形式をとる。ここで我々は、話し手が G_i から G_{i+1} へ物体群を絞り込むとき、群 G_i と G_{i+1} の規模 (dimension) の間に最適な比が存在すると仮定する。例えば非常に大きな群から非常に小さな群に絞り込むことは、聞き手を混乱させやすいと予想でき、そのような場合には間に中間的な大きさの群を挟んだ表現がより適切であると考えられる。例えば、次の 2 つの表現を考える。この 2 表現はともに図 2 中の物体 x を指している。表現 (ii) は表現 (i) よりも単純であるが、聞き手は表現 (ii) よりも表現 (i) を好むだろう。

- (i) 「左奥の 3 つの玉のうち一番右の玉」
- (ii) 「右から 4 番目の玉」

実際、2.1 節の実験 1 で収集したデータは表現 (i) を含んでいたが、表現 (ii) はデータ中に存在しなかった。また、提案アルゴリズムは (i) と (ii) の両表現を生成し、それぞれの表現の得た評価値は 44.4 と 32.1 であった。我々の仮定が正しければ、群間の適切な規模比をもつ表現を選択することで、よりよい表現を選択することが期待できる。

計算式

SOG のスコアの計算式を (6) に示す。SOG 全体のスコアは、SOG 内で隣接する群間に対して関数 f_1 と f_2 が与えるスコアを平均することで求める。 n は SOG が含む群の数である。

$$score(SOG) = \frac{1}{n-1} \left\{ \sum_{i=0}^{n-3} f_1 \left(\frac{dim(G_{i+1})}{dim(G_i)} \right) + f_2 \left(\frac{dim(\{x\})}{dim(G_{n-2})} \right) \right\} \quad (6)$$

群 G の規模 $dim(G)$ は、群の幾何中心から群の各メンバまでの距離の平均として表す。 $\{x\}$ の規模は定数とする。この $\{x\}$ の特異性のために、 f_2 を f_1 から分離して定義する。ただし、両関数の基本的な概念は同一である。

二つの群間の最適な比率 f_1 および f_2 は、2.2 節に記した実験で集めたデータを二次回帰分析することで求めた。 f_1 および f_2 は、規模比と式 (1) で与えられる評価値の間の相関関係を表す回帰曲線である。規模比と正答率の間には直接的な相関関係を見いだせなかった。

4.2 結果

生成した表現に対して式 (6) が与えるスコアがどの程度人間の評価（式 (1)）と一致するかを調べた．一致度は 3.3 節で説明した 20 布置を用いて計算した．

はじめに，式 (6) が与えるスコアに従って，生成した表現を順序付けた．同様に式 (1) が与える評価値に従って，別に順序付けた．こうして順序づけられた二つのリストから，全ての 2 表現間の順序関係を抽出した．一致度は，全ての 2 表現間の順序関係のうち，二つのリスト間で同じ順序関係にあるものの割合として定義した．

式 (6) によるスコアと式 (1) による評価値との間の一致度は 45.8%と低く，スコアは SOG の評価値を十分に予測できなかった．

以上のことから，SOG に含まれる群の間の規模比だけをもとに最終的な表現の良し悪しを判断することはできないことがわかった．複数の表現から適切な表現を選択するためには，他の要素を考慮しなくてはならない．

5 おわりに

本論文は，知覚的群化に基づいて多項間関係を用いる参照表現を生成する手法を提案した．提案アルゴリズムは，心理学実験によって収集した表現に対する分析に基づいて構築し，23 人の被験者によって評価した．アルゴリズムが生成した表現はその 89.2%で対象物体の特定に成功し，人間の生成した参照表現とほぼ同等の評価を得た．

今後の課題として，まず位置関係以外の関係の利用があげられる．本論文では，群間の位置的な関係だけに焦点を当てた．しかし，色の度合いや大きさなど他の種類の関係も同様に扱われるべきである．Thórisson のアルゴリズムは，物体間の類似性を特徴別に定義することで，これらの関係も位置関係と同様に扱うことができる．しかしながら，複数の種類の関係を同時に用いる場合，提案アルゴリズムのステップ 1 で生成される群同士が交差しないという前提が必ずしも成立しなくなる．そのため，ステップ 2 の SOG 生成手法を再考する必要がある．

もう一つの課題として提案手法と従来手法との融合がある．本論文では，物体の固有の属性がまったく対象特定の役に立たない状況に限定して焦点を当てた．しかしそのような状況は一般的ではない．従来属性ベースの手法と提案手法を融合した生成アルゴリズムが必要である．また，外部参照関係も使えるようにする必要がある．

考えられる一つの方向性は，Krahmer らの提案する手法 (Krahmer et al. 2003) の拡張である．彼らの手法は，物体の配置をラベル付き有効グラフとして記述し，表現内容の選択をユニークでコスト最小の部分グラフを発見するグラフ探索問題と見なす．探索はコスト関数によって主に制御される．知覚的群化によって認識された群を，図 7 に示すように通常の物体と同様に見なせば，彼らの手法を適用することができる．これによって，内部参照関係（図 7 中で front, middle, back などのラベルがついた辺）だけでなく，外部参照関係（例えばノード Group 1 からノード Table への辺）も扱うことができる．しかしながら，群をグラフ表現の中に導入する

ことは探索を制御するコスト関数の設計を難しくする．簡潔でわかりやすい表現を生成するためには適切に設計されたコスト関数が必要である．そうでなければ，図 1 に対して，「玉の前の玉の前の玉」などといった誤りではないが適切ではない表現を生成してしまう．

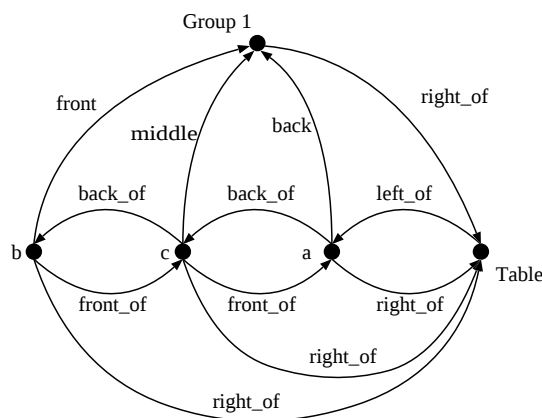


図 7: 図 1 中の状況をあらわすグラフ

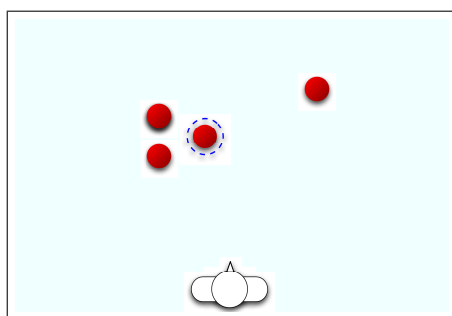
参考文献

- Appelt, D. E. (1985). “Planning English Referring Expressions.” *Artificial Intelligence*, **26**, 1–33.
- Cavazza, M., Charles, F., and Mead, S. J. (2002). “Character-based Interactive Storytelling.” *IEEE Intelligent Systems*, **17** (4), 17–24.
- Clark, H. H. (1973). “Space, time, semantics, and the child.” In Moore, T. E. (Ed.), *Cognitive development and the acquisition of language*, pp. 65–110. Academic Press.
- Dale, R. (1992). “Generating Referring Expressions: Constructing Descriptions in a Domain of Objects and Processes.”. MIT Press, Cambridge.
- Dale, R. and Haddock, N. (1991). “Generating referring expressions involving relations.” In *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics (EACL’91)*, pp. 161–166.
- Dale, R. and Reiter, E. (1995). “Computational interpretations of the Gricean maxims in the generation of referring expressions.” *Cognitive Science*, **19** (2), 233–263.
- Heeman, P. and Hirst, G. (1995). “Collaborating referring expressions.” *Computational Linguistics*, **21** (3), 351–382.
- Herskovits, A. (1986). *Language and Spatial cognition: an interdisciplinary study of the prepositions in English*. Cambridge University Press.

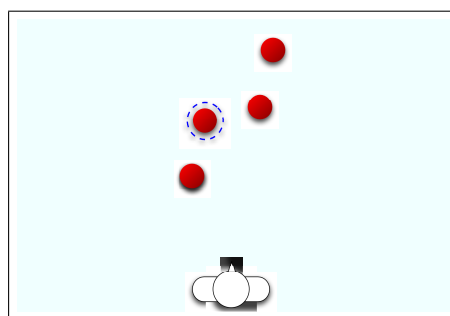
- Horacek, H. (1997). "An Algorithm for Generating Referential Descriptions with Flexible Interfaces." In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pp. 206–213.
- Jörding, T. and Wachsmuth, I. (2002). "An Anthropomorphic Agent for the Use of Spatial Language." In Coventry, K. R. and Olivier, P. (Eds.), *Spatial Language*, pp. 69–85. Kluwer Academic Publishers.
- Krahmer, E. and Theune, M. (2002). "Efficient context-sensitive generation of descriptions." In van Deemter, I. K. and Kibble, R. (Eds.), *Information Sharing: Givenness and Newness in Language Processing*, pp. 223–264. CSLI Publications, Stanford, California.
- Krahmer, E., van Erk, S., and Verleg, A. (2003). "Graph-Based Generation of Referring Expressions." *Computational Linguistics*, **29** (1), 53–72.
- Levinson, S. C. (Ed.) (2003). *Space in Language and Cognition*. Cambridge University Press.
- Tanaka, H., Tokunaga, T., and Shinyama, Y. (2004). "Animated Agents Capable of Understanding Natural Language and Performing Actions." In Prendinger, H. and Ishizuka, M. (Eds.), *Life-Like Characters*, pp. 429–444. Springer.
- Thórisson, K. R. (1994). "Simulated Perceptual Grouping: An Application to Human-Computer Interaction." In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pp. 876–881.
- van Deemter, K. (2002). "Generating referring expressions: Boolean extensions of the incremental algorithm." *Computational Linguistics*, **28** (1), 37–52.
- van der Sluis, I. and Krahmer, E. (2000). "Generating Referring Expressions in a Multimodal Context: An empirically oriented approach." Presented at the CLIN meeting 2000, Tilburg.

A 実験に用いた布置

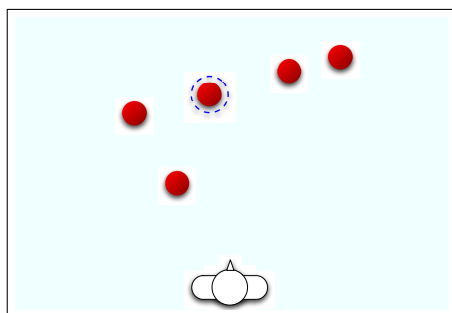
布置中の玉の数は最低 3 個，最大 9 個である．布置の作成は，最初に玉を 3 個含む布置を作成し，そこに玉を一つずつ付け加えていった．元となる三つの布置から 18 布置を作成し，そこから 12 の布置を選択した．3 番，10 番，6 番の布置は 7 番の布置を元に作成した．1 番，11 番，4 番，8 番の布置は 9 番の布置を元に作成した．2 番，12 番の布置は 5 番の布置を元に作成した．



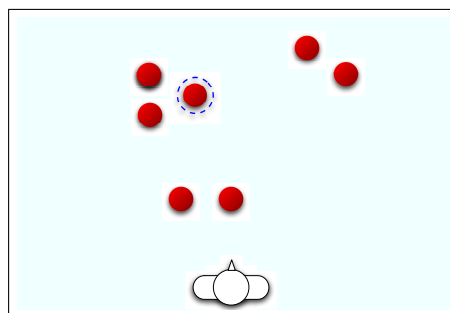
1



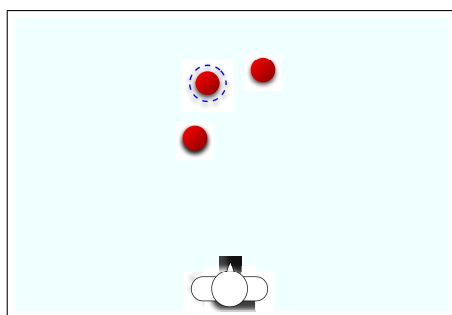
2



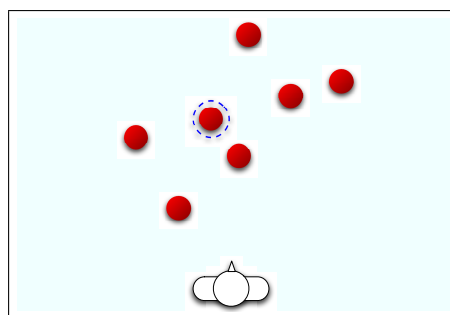
3



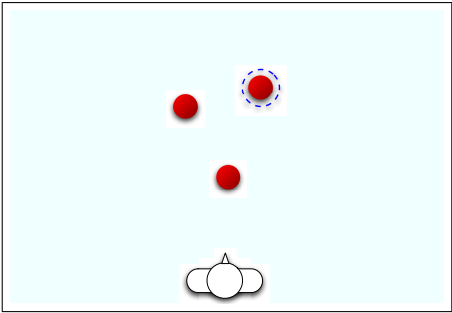
4



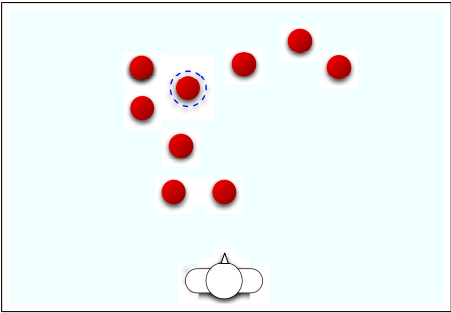
5



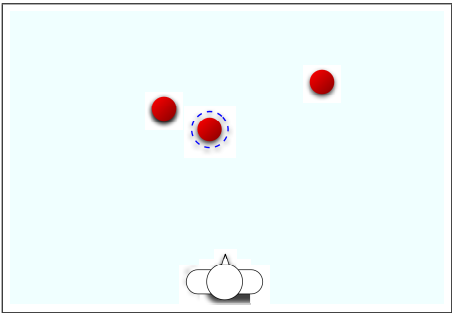
6



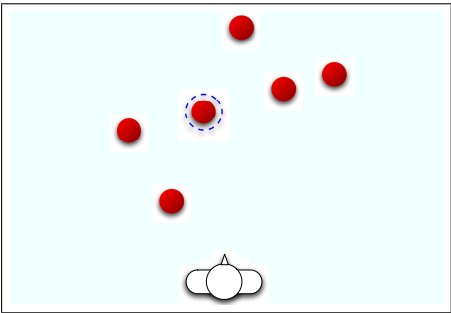
7



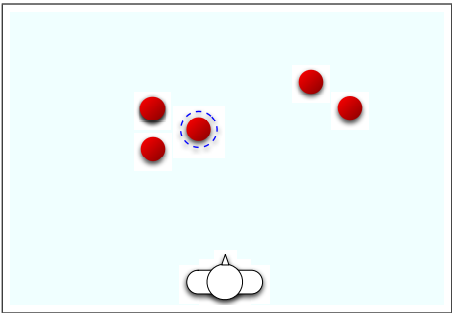
8



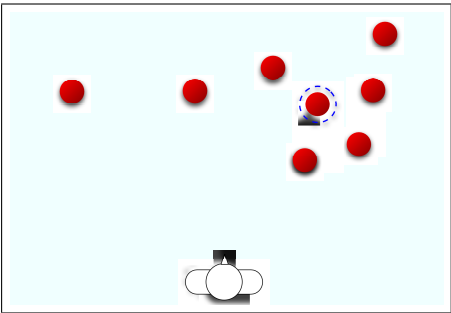
9



10



11



12

略歴

船越 孝太郎 (正会員): 2002 年東京工業大学大学院情報理工学研究科修士課程修了。2005 年同大学同研究科博士課程修了。現在, 同大学同研究科特別研究員。博士 (工学)。自然言語理解, 音声対話に関する研究に従事。情報処理学会, 人工知能学会, IEEE, AAAI 各会員。

渡辺 聖 (非会員): 2004 年東京工業大学工学部情報工学科卒業。同年同大学院情報理工学研究科修士課程進学, 在学中。自然言語生成に関する研究に従事。

栗山 直子 (非会員): 1998 年東京工業大学大学院社会理工学研究科修士課程修了。2002 年同大学大学院博士課程修了。現在, 同大学大学院社会理工学研究科助手。博士 (学術)。認知科学会, 認知心理学会, 日本心理学会, 教育心理学会, 教育工学会, 各会員。

徳永 健伸 (正会員): 1961 年生。1983 年東京工業大学工学部情報工学科卒業。1985 年同大学院理工学研究科修士課程修了。同年 (株) 三菱総合研究所入社。1986 年東京工業大学大学院博士課程入学。現在, 同大学大学院情報理工学研究科助教授。博士 (工学)。自然言語処理, 計算言語学, 情報検索などの研究に従事。情報処理学会, 認知科学会, 人工知能学会, 計量国語学会, Association for Computational Linguistics, ACM SIGIR 各会員。

(2005 年 9 月 13 日 受付)

(2005 年 12 月 16 日 再受付)

(2005 年 12 月 28 日 再々受付)

(2005 年 12 月 30 日 採録)