

# 知覚的群化を利用した参照表現の生成

渡辺聖

船越孝太郎

徳永健伸

東京工業大学 大学院情報理工学研究科 計算工学専攻

{satoru\_w, koh, take}@cl.cs.titech.ac.jp

## 1 序論

参照表現とは対象物体を他の物体と混同せずに特定する言語表現であり、人間同士、あるいは人間とコンピュータの間のコミュニケーションにおいて重要な役割を果たす [1]. 参照表現には談話文脈 (例: 「さっきの～」) やパラ言語情報 (話し手の視線や指差し動作) などを利用する表現も含まれるが、本論文では対象物体固有の属性や他の物体との関係のみを用いる参照表現の生成を研究対象とする。(例: 「ドアの前にある赤い椅子」)

従来の参照表現の生成手法 [2-4, 7, 8] は、対象物体固有の属性や他の物体との2項関係のみを用いることが多かった。そのため、指示すべき物体とその他の物体との間に有意な外見的特徴の差がなく、他の物体との位置関係も弁別の用を成さないような状況では、適切な参照表現を生成することができない。

例として図1の状況を考える。属性や2項関係のみを利用する従来のアルゴリズムでは、物体  $b$  を示すために「玉の手前の玉で玉の奥の玉」といった不自然な表現を生成してしまうか、参照表現の生成は失敗となっていた。ここで「真ん中の玉」のような自然な表現を生成するためには、知覚的に特徴のある物体群を認識し、その群の中での  $n$  項関係を利用する必要がある。

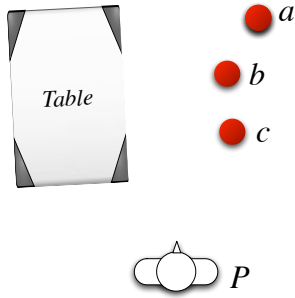


図1: 従来手法での表現生成が困難な例

この問題を解決するために我々は、知覚的群化を行ない、物体群の間関係を用いて参照表現を生成する手法を提案した [5, 6, 11, 12]. 知覚的群化 [9] とは外見的特徴の差異の少ない物体や相互に近接した物体を1つの群として認識することである。この手法によって物体の  $n$  項関係を利用した参照表現の生成が可能となったが、この手法の想定するドメインは、同形同色同大の物体を複数配置した2次元空間という非常に限られたも

のであったため、一般的な状況には対応できなかった。

本論文では、[5, 6, 11, 12] で提案した手法を拡張し、従来より利用されてきた色、形、大きさ等の様々な属性情報を利用可能とする参照表現の生成手法を提案する。

以降、2節で知覚的群化を利用する参照表現の生成において中心的な役割を担う SOG (Sequence Of Groups) について説明する。3節では参照表現の生成手法を提案する。そして4節で、提案手法の評価と考察を行い、5節でまとめとする。

## 2 SOG (Sequence Of Groups)

我々は、参照する空間の状況と参照表現との間の中間表現として、SOG を提案した [5, 6, 11, 12]. これは、物体全体の群  $G_0$  から始まり対象物体のみを含む群  $G_n$  に至る物体群の列である。ただし隣接する物体群の間にはそれらを結ぶ関係を表す記号  $R_i$  が挿入される。主要部後置型である日本語において、人は物体全体から比較的小さな物体群へ参照範囲を絞り込みながら、あるいは他の物体を参照しながら対象物体を特定する。SOG はこの対象物体特定の過程を抽象化したものであり、以下のように定式化できる。

$$\text{SOG: } [G_0 R_0 G_1 R_1 \dots G_i R_i \dots G_{n-1} R_{n-1} G_n]$$

$R_i$  は群  $G_i$  と群  $G_{i+1}$  を結ぶ関係であるが、関係には内部参照関係と外部参照関係の2種類がある。

**内部参照関係**  $G_i \supset G_{i+1}$  の包含関係となる。この関係は絞り込みに利用する素性の種類だけ存在する。ここで素性は物体固有の属性情報と物体間関係 (近接性等) の両方を含むものとする。以下に本論文が対象とする素性を表す記号とその説明を示す。

- > type > : タイプを利用した絞り込み
- > shape > : 形を利用した絞り込み
- > color > : 色を利用した絞り込み
- > size > : 大きさを利用した絞り込み
- > space > : 空間的 (位置的) 絞り込み

**外部参照関係**  $G_i \cap G_{i+1} = \phi$  となる。この関係は空間的な関係に限られる。

- : 空間的 (位置的) 関係

ここで図2に示す空間に対して、参照表現とその参照表現に対応する SOG の例を示す。全体群は  $\{all\}$  と略記する。また対象物体を  $b1$  とし、図2中に破線で囲んで示す。

**参照表現:** 「四角い<sub>(1)</sub> 机<sub>(2)</sub> の 左<sub>(3)</sub> の 小さい<sub>(4)</sub> 玉<sub>(5)</sub>」

$$\text{SOG} : [ \{all\} > type > \{t1, t2, t3\} \geq shape \geq_{(1)} \{t1\}_{(2)} \Rightarrow_{(3)} \{b1, b2, b3\} \geq size \geq_{(4)} \{b1\}_{(5)} ]$$

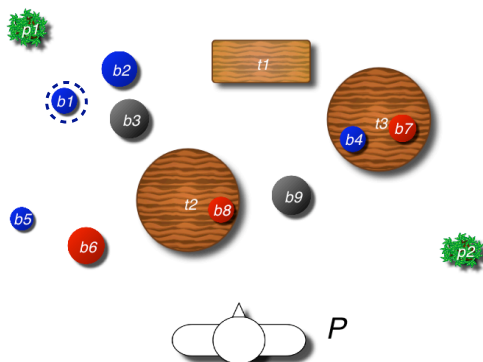


図 2: 参照する空間の例

### 3 参照表現の生成手法

本論文が提案する参照表現の生成手法は (1) 知覚的群化 (2) SOG 生成 (3) 表現の付与 (4) 順位付けの 4 ステップから成る。以降、これらのステップをそれぞれ説明する。

#### (1) 知覚的群化

知覚的群化のために以下の素性を使う。

1. 物体のタイプ
2. 物体の形
3. 物体の色
4. 物体の大きさ
5. 物体間の空間的関係

1. 物体のタイプは特に重要視される。なぜなら、人は一般的に異なるタイプの物体を同一の物体群として捕らえることは少なく、また『タイプ』の属性はその物体を最も単純に表現できるからである。これは Incremental Algorithm [4] でも論じられている。

本手法では、まず物体のタイプを利用した群化を行ない、それぞれのタイプの物体群に対して 2.~5. の素性を利用した群化を行なう。本手法ではそれぞれの素性を単独で Thórisson が提案した知覚的群化の手法 [9] に適用する。

生成した各物体群には、知覚的群化の際に利用した素性の情報を付与しておく。本手法の対象とするドメインでは、{全体 (*all*), タイプ (*type*), 形 (*shape*), 色 (*color*), 大きさ (*size*), 空間的屬性 (*space*)} の 6 つの素性を定義した。

例として図 2 の状況に対して知覚的群化を行なった結果生成される物体群を以下に示す。

全体群 ( <i>all</i> )	{ <i>t1, t2, t3, p1, p2, b1, b2, ..., b9</i> }
単体群 ( <i>space</i> )	{ <i>t1</i> }, { <i>t2</i> }, { <i>t3</i> }, { <i>p1</i> }, { <i>p2</i> }, { <i>b1</i> }, { <i>b2</i> }, ..., { <i>b9</i> }
タイプ ( <i>type</i> )	{ <i>t1, t2, t3</i> }, { <i>p1, p2</i> }, { <i>b1, b2, ..., b9</i> }
形 ( <i>shape</i> )	{ <i>t1</i> }, { <i>t2, t3</i> }
色 ( <i>color</i> )	{ <i>b1, b2, b4, b5</i> }, { <i>b3, b9</i> }, { <i>b6, b7, b8</i> }
大きさ ( <i>size</i> )	{ <i>b1, b4, b5, b8</i> }, { <i>b2, b3, b6, b7, b9</i> }
近接性 ( <i>space</i> )	{ <i>t1, t3</i> }, { <i>b1, b2, b3</i> }, { <i>b4, b7, b8, b9</i> }, { <i>b5, b6</i> }
閉包 ( <i>space</i> )	{ <i>b4, b7</i> }, { <i>b8</i> }

ここで閉包とは、他の物体または領域に囲まれている物体に対する群化である。この例の場合、机の上に乗っている玉をそれぞれ群化する。

#### (2) SOG 生成

ここでは深さ優先探索を利用して、考えられる全ての SOG を生成する。SOG 生成の手順を以下に示す。

(2.1) SOG として空リストを生成し、第 1 要素に全体群を追加する。

(2.2) 現在の SOG の最後の要素  $G_i$  に対して場合分けをし、それぞれの処理を行なう。

1.  $G_i$  が全体群の場合：
 

空間内に存在する物体のタイプの中から対象物体より顕現性<sup>1</sup>が高いか対象物体と同じタイプを選択し、群間の関係 “ $> type >$ ” とそのタイプの全体群を SOG に追加する。選択するタイプがない場合、探索を終了する。
2.  $G_i$  が対象物体以外の複数の物体を含む群の場合：
 

(1) で生成した群の中から、 $G_i$  に包含される物体群を 1 つ選択し、選択した群に付与された素性情報から群間の関係  $R_i$  を生成する<sup>2</sup>。この  $R_i$  と選択した群を SOG に追加する。選択する群がない場合、バックトラックをする。
3.  $G_i$  が対象物体以外の単体からなる場合：
 

(1) で生成した群の中から、対象物体を含む群で、 $G_i$  に含まれる物体 (参照物体) から適切な位置関係にある物体群を選択する。ここで適切な位置関係とは、以下の条件 (a), (b), (c) を満たすものである。

(a) 対象物体を含む群の全要素が、参照物体から「奥, 右奥, 右, 右手前, 手前, 左手前, 左, 左奥」<sup>3</sup>のいずれかの表現で表せる同一の方向にあるか、参照物体の「上」にある。

<sup>1</sup>本論文のドメインでは机, 木, 玉の順に顕現性が高いと定義した

<sup>2</sup>素性情報の記号の前後に “ $>$ ” を付けたものを  $R_i$  とする

<sup>3</sup>参照物体の重心を原点とし、空間内の人物が向いている方向を「奥」の方向として方向を 8 等分し、この順に方向を割り当てる。

- (b) (a) と同一の方向で、かつ対象物体群のいずれの要素よりも参照物体から近い位置に同じタイプの他の物体が存在しない
- (c) 対象物体群はいずれかのタイプの全体群ではない

選択した物体群の素性情報から  $R_i$  を生成し、この  $R_i$  と選択した群を SOG に追加する。選択する群がない場合、バックトラックをする。

4.  $G_i$  が全体群ではなく、対象物体を含む複数の物体を含む群の場合：
- (1) で生成した群と  $G_i$  との積集合をとった群を生成し、付与されている素性情報もコピーする。生成した群が対象物体を含むならば、付与された素性情報から群間の関係  $R_i$  を生成し、 $R_i$  と選択された群を SOG に追加する。群を生成できない場合、バックトラックをする。
5.  $G_i$  が対象物体のみの単体群の場合：
- この場合 SOG は完成しているので、この SOG を記録する。バックトラックを行ない、他の SOG を探索する。

(2.3) 再び (2.2) へ戻る。

図 2 の状況に対して生成した SOG の一部を示す。

- $\{\{all\} > type > \{t1, t2, t3\} > shape > \{t1\} \rightarrow \{b1, b2, b3\} > size > \{b1\}\}$
- $\{\{all\} > type > \{p1, p2\} > space > \{p1\} \rightarrow \{b1\}\}$
- $\{\{all\} > type > \{b1, b2, \dots, b9\} > space > \{b1, b2, b3\} > color > \{b1, b2\} > space > \{b1\}\}$

### (3) 表現付与

SOG の各要素に表現を付与して参照表現を生成する。表現の付与には以下の規則を用いる。規則 1.1, 1.2 は物体群について、規則 2.1, 2.2 は群間の関係についての規則である。

**規則 1.1** 全体群 ( $\{all\}$ ) と各タイプの全体群は言語化しない。ただし単体群の場合を除く。2 節の例からもわかるように、一般的に全体群は省略される。

**規則 1.2** 各物体群に『個数+タイプ名+「の(うち)」』を付与する。ただし物体群の後に “> shape >”, “> color >”, “> size >” が続く場合は言語化しない。これは、空間的な関係を利用することなく物体群を絞り込む場合、絞り込みの対象となる群の表現は省略し、絞り込みに利用した属性のみを言語化するのが自然である [4] ためである。逆に空間的關係を利用した関係が続く場合、参照する物体群を言語化することで曖昧性のない参照表現を生成できる。ここで、“→” が続く場合には「の」を、“> space >” が続く場合

には「のうち」を付与する。また単体群に対しては、「1 つの」という様な個数の表現は省略する。

**規則 2.1** タイプの絞り込み (> type >) は言語化しない。タイプ絞り込みはその関係を言語化せず、物体群に対して『タイプ名』を付与することでその役目を果たす。

**規則 2.2** 各関係には『関係が示す情報』を付与する。“> shape >”, “> color >”, “> size >” にはそれぞれの属性の値の表現を付与し、“→”, “> space >” に対しては、前後の群の各要素の位置座標を計算し、適切な空間表現を付与する。

(2) の SOG の例に付与した表現を以下に示す。

- $\{\{all\} > type > \{t1, t2, t3\} > shape > \{t1\} \rightarrow \{b1, b2, b3\} > size > \{b1\}\}$   
「四角い 机の 左の 小さい 玉」
- $\{\{all\} > type > \{p1, p2\} > space > \{p1\} \rightarrow \{b1\}\}$   
「左の 木の 手前の 玉」
- $\{\{all\} > type > \{b1, b2, \dots, b9\} > space > \{b1, b2, b3\} > color > \{b1, b2\} > space > \{b1\}\}$   
「左奥の 青い 2 つの 玉のうち 左の 玉」

### (4) 順位付け

出力表現を決定するための順位付けの手順 (4.1)~(4.4) を以下に示す。コストが小さいものほど適切な表現となる。

(4.1) SOG 内の各関係に以下の要領で [0,1] のコストを付ける。“> type >”, “> color >”, “> size >” の優先度は Incremental Algorithm [4] を参考にした<sup>4</sup>。

> type > 言語化しないので未定義とする

> shape > 0.2

> color > 0.4

> size > 「大きい」0.6 「小さい」0.8 「中くらいの」1.0

> space >, → ポテンシャル関数 [10] を参考に人手で計算式を作った。ただし本論文が対象とするドメインでは、「~の上の」という表現は顕現性が特に高いと考えられるのでコストを 0 とした。

(4.2) (4.1) で計算した各関係のコストの平均値を  $cost_{rel}$  とする。

(4.3) 表現の簡潔さのコストを、 $cost_{len} = \frac{\text{(文字数)}}{\text{(全表現の最大文字数)}}$  として計算する。

(4.4) 総合コストを  $cost = \alpha \times cost_{rel} + (1 - \alpha) \times cost_{len}$  として計算する。ただし  $\alpha$  は重み付けの定数値であり、提案手法では  $\alpha = 0.5$  とした。

<sup>4</sup>人間は絶対的な素性(色, 形など)を相対的な素性(大きさなど)よりも優先する傾向がある。

(3)の参照表現の例に対して計算したコストを、表現全体(40個)の中での順位と共に以下に示す。数値は総合コスト[関係のコスト, 簡潔さのコスト]である。

左の木の手前の玉 : 0.25 [0.21,0.29] 1位

四角い机の左の小さい玉 : 0.46 [0.52,0.39] 3位

左奥の青い2つの玉のうち左の玉 : 0.53 [0.53,0.54] 10位

## 4 考察

提案手法を評価するため、上記のアルゴリズムを実装し、大学生18人を対象に心理実験を課題1と課題2に分けて行なった。被験者は課題1の後に課題2を行なったが、それらの関連については知らせていない。実験に使用した布置は、物体の配置も個数もランダムに決定したもので、対象物体もランダムに選んだ。ただし布置は、参照表現が5個以上生成され、かつ「机」のような1単語の表現では対象物体を特定できないような、対象物体と同じタイプの物体が2つ以上存在するものに限定した。

**課題1** 被験者に、布置の画像とその布置に対して生成した参照表現のうち最もスコアが高かったものを示し、その参照表現の指し示す物体を選ばせた。この方法で、それぞれの被験者に用意した20個の布置に対する参照表現を評価させた。

この結果、対象物体特定の正解率は全体で95%(342/360)だった。このことから、実装システムが生成した参照表現は高い対象物体特定能力をもつと言える。

**課題2** 対象物体を明示した布置の画像と、その布置に対して生成した参照表現のうち上位5つを被験者に示し、その中で被験者が最もよいと思う表現を選ばせた。判断基準は明確に設定せずに、最大限各自の判断に任せた。布置は課題1と同じもの20個を使用した。

この結果を表1に示す。表現は1~5の順にスコアが高かったものである。上位2位の参照表現の得票率が全体の72%を占めている。このことから本論文で提案した参照表現に対するスコア付けの手法が有効であるといえる。

表1: 実験結果: 得票率

表現	1	2	3	4	5	計
度数	134	125	59	22	20	360
得票率	0.37	0.35	0.16	0.06	0.06	1.00

## 5 まとめ

本論文では知覚的群化を利用した参照表現の生成手法を提案した。また参照表現と参照する空間の状況との間にSOGという概念を提案し、これを利用した。

本論文の提案する手法は以前に提案した群間の内部参照関係のみを用いた手法を拡張し、より一般的など

メインに対応できるようにしたものである。提案手法は(1)知覚的群化, (2)SOG生成, (3)表現付与, (4)順位付けの4ステップから成る。この手法を実装したシステムが生成した参照表現の対象物体特定の正解率は95%であった。またスコア付けによる参照表現の順位付けは人間の感覚に近いことがわかった。

提案手法のスコア付けの手法はIncremental Algorithmを参考に人手で作成したものであり、評価実験課題2で高い評価を得たが、この結果は実装システムの生成した参照表現が人間の生成する参照表現にどこまで近づけたかを示すものではない。人間が生成する参照表現との比較は今後の課題である。

## 参考文献

- [1] Donna K. Byron. Understanding referring expressions in situated language: Some challenges for real-world agents. In *the First International Workshop on Language Understanding and Agents for the Real World*, 2003.
- [2] Robert Dale. Generating referring expressions: Constructing descriptions in a domain of objects and processes, 1992. MIT Press, Cambridge.
- [3] Robert Dale and Nicholas Haddock. 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, 1991.
- [4] Robert Dale and Ehud Reiter. Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, Vol. 19, No. 2, pp. 233-263, 1995.
- [5] Kotaro Funakoshi, Satoru Watanabe, Naoko Kuriyama, and Takenobu Tokunaga. Generating referring expressions using perceptual groups. In *Third International Conference on Natural Language Generation: INLG04*, Jul 2004.
- [6] Kotaro Funakoshi, Satoru Watanabe, Naoko Kuriyama, and Takenobu Tokunaga. Generation of relative referring expressions based on perceptual grouping. In *the 20th International Conference on Computational Linguistics*, 2004.
- [7] Emiel Krahmer and Mariët Theune. Efficient context-sensitive generation of descriptions, 2002. In Kees van Deemter and Rodger Kibble, editors, *Information Sharing: Givenness and Newness in Language Processing*. CSLI Publications, Stanford, California.
- [8] Emiel Krahmer, Sebastiaan van Erk, and André Verleg. Graph-based generation of referring expressions. *Computational Linguistics*, Vol. 29, No. 1, pp. 53-72, 2003.
- [9] K. R Thórisson. Simulated perceptual grouping: An application to human-computer interaction. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pp. 876-881, 1994.
- [10] Takenobu Tokunaga, Tomofumi Koyama, and Suguru Saito. Meaning of Japanese spatial nouns. In *Proceedings of the Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics: Formalisms and Applications*, pp. 93-100, 2005.
- [11] 渡辺聖. 集合内の関係を用いた参照表現の生成に関する研究. 学士論文, 東京工業大学, 2004.
- [12] 渡辺聖, 船越孝太郎, 栗山直子, 徳永健伸. 集合内の関係を用いた参照表現の生成. 言語処理学会第10回年次大会, pp. 516-519, 3 2004.