

集合内の関係を用いた参照表現の生成

渡辺聖*

船越孝太郎*

栗山直子**

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

** 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
kuriyama@hum.titech.ac.jp

1 序論

参照表現の生成は自然言語処理の重要な課題の1つであり、多くの研究がなされてきた [1-5].

言語における物体の指示は、以下のような情報(素性)を組み合わせて、対象物体を特定可能な表現を生成する事で行なわれる.

1. 対象物体固有の属性
(例:「赤い玉」,「日本製のテレビ」)
2. 他の物体との関係
(例:「ドアの前の椅子」,「小さい方のグラス」)
3. 対象物体に関する聞き手の信念
(例:「君の好きな本」)
4. 談話文脈
(例:「さっきの場所」)
5. パラ言語情報
(例:話し手の視線や指差し動作)

従来の参照表現生成手法 [1-5] は、対象物体固有の属性や他の物体との二項間関係を素性として用いることが多い. そのため、指示すべき物体とその他の物体との間に外見的特徴の有意な差がなく、他の物体との位置関係も弁別の役に立たない状況では、適切な参照表現¹を生成することができない.

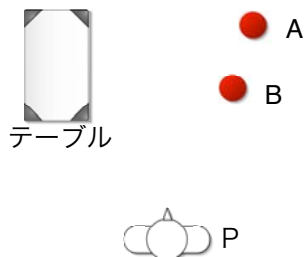


図 1: 従来の手法では参照表現が生成できない例

例として、図 1 の状況で図中の人物 P に玉 B を指し示すことを考える. ここで、人物 P は「A」や「B」のようなラベル情報は、話し手と共有していないとする.

¹適切な参照表現とは、聞き手が話し手の指す物体を、他の物体と混同することなく特定できる表現を意味する. ただし、極度に不自然な表現や冗長な表現はその範囲に含まない.

玉 B は外見的特徴だけでは玉 A との区別がつかない. そこで玉 B とテーブルとの位置関係を利用したいが、玉 A, B は共に「テーブルの右」という位置関係となり玉 B を特定するには不十分である². そこで玉 A との位置関係を利用し「(玉 A)の手前」と表現したいが、それは玉 A の参照表現生成の問題へと帰着する. しかし、玉 A は玉 B と同様に、固有属性及び他の物体との二項関係だけでは一意に特定できず表現生成は不可能となる.

van der Sluis & Krahermer は、類似した対象が多数存在し言語的に弁別しがたい場合には、指さしなどを使う事で解決する事を提案している [7]. しかし、話し手と聞き手の位置関係によっては指さしや視線は必ずしも有効ではない.

図 1 の状況においては、話し手は「手前の玉」というごく簡単な表現を用いる事で、パラ言語情報を利用しなくても玉 B を人物 P に指し示す事ができる. このような表現を生成するためには、玉 A と玉 B からなる集合を認識し(顕在性の高いまとまりの認識)、その集合内での関係を用いて対象を表現する能力(二項間以上の関係を使った参照表現の生成)が必要となる³.

そこで本論文では、類似した物体の集合を認識し、その集合内での関係を用いた参照表現の生成手法を提案する. 提案手法は、Thórisson [6] が提案する手法を利用して集合を認識し、その出力を用いて参照表現を生成する.

集合内の関係には位置関係以外にも、大きさや色などの固有属性の程度の差を表す関係⁴も存在するが、今回は位置関係のみを対象とする. 聞き手の信念や、談話文脈も利用しない.

また、一つの物体を表現する過程で、往々にして複数の集合が参照される. このような場合、集合間には参照関係⁵と包含関係⁶の2種類があるが、今回は包含関係のみ用いる. 従って、単独の物体から集合を参照する表現⁷も本手法では扱わない.

以降、2 節では、集合内の関係を用いた参照表現を収集するために行なった実験とその結果について説明す

²参照枠の解決は適切に行なえるものと仮定し、本論文では取り扱わない

³Krahermer et al. [5] は彼らの手法が n 項関係にも適用できると述べているが、詳細は述べておらず、図 1 の様な状況をそのまま扱えるとは考えられない

⁴例:「2 番目に大きな玉」,「最も明るい色の玉」

⁵例:「右の 5 つの玉のすぐ奥にある 3 つの玉」

⁶例:「右の 5 つの玉の中の奥の 3 つの玉」

⁷例:「テレビの前の二つ」

る。そして、3節で、提案手法について述べる。4節では提案手法について考察を行なう。

2 参照表現の採集と分析

集合内の関係を用いた参照表現のパターンを集めるために、大学生42人を対象に採集実験を行なった。また得られた表現の妥当性を調べるために、異なる大学生44人を対象に評価実験を行なった。

2.1 表現採集実験

実験手法 同形・同大・同色の物体を複数個配置した2次元画像を被験者に示し、破線で囲まれた物体を画像中に描かれた第三者に伝える場合、どのように表現するかを尋ねた。図2は被験者に提示した画像の一例である⁸。物体の布置は12種類用意し、被験者全員にそれぞれの布置について指定された対象物体を表現させた。各布置において指定する物体は1つに固定した。被験者が作る表現は「～を取ってください」という文に限定し、表現不可能と考える場合には放棄する事も許した。

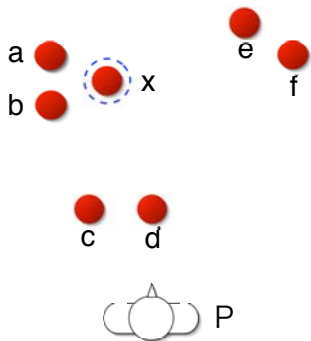


図 2: 表現採集に使った視覚刺激の例

分析 学生42人から得られた表現の総和は476表現で、放棄された数は28であった。収集した表現から、被験者がどのように物体の集合を認識し利用しているかを判断し、それぞれの表現に集合の認識と絞り込みの過程を示すタグを与えた。ここでタグとは、話者が表現生成に使用した集合を順に並べたものである。

以下に図2に対して実際に得られた表現の一例とそれに与えたタグを示す。

例: 「左奥にある3つの赤い玉のうち一番右の赤い玉」
 $\rightarrow \{a, b, c, d, e, f, x\}, \{a, b, x\}, \{x\}$

$\{a, b, c, d, e, f, x\}$: 全ての玉の集合

$\{a, b, x\}$: 左奥の3つの玉

$\{x\}$: 指示すべき玉

⁸あらかじめ物体に割り当てた標識 $a \sim f, x$ は被験者には見えない。

絞り込みは必ず全体集合から始まるので、タグは $\{\{\text{全ての物体}\}, \dots, \{\text{対象物体}\}\}$ という形となる。このようにタグを与えることによって、表現を抽象化し分類することができる。実際に、得られた表現は1布置につき42個程あるが、平均8.4種類のタグに分類できた。また逆に表現を生成する際には、まずタグを生成し、そのタグに表現を付与することで分類に即した表現の生成が可能となる。

集合間の関係には包含関係と参照関係があるが、包含関係のみを用いた表現は採集した表現の約8割を占めていた。1節で述べたように本論文は包含関係のみを用いた表現のみを扱う。

2.2 表現の評価実験

実験手法 採集実験で用いた画像とその画像から得られた表現の1つを被験者に示し、与えられた表現がどの玉を指しているのかを答えさせた。同時に自分の解答の確信度、与えられた表現の簡潔さと自然さをそれぞれ8件法で評価させた。

評価対象には、先の実験で得られたものの中から代表的なものを1布置につき最大10個、全部で117表現を選んだ⁹。これを学生44人に、1人につき約32個の表現を評価させた。

分析 有効な評価の総数は1429で、1つの表現に対して12~13個の評価が得られた。表現の適切性の評価には以下の式を用いた。

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

この評価値が高いものはより適切な表現であると考え、表現生成の参考とした¹⁰。今回の分析では、集合間の参照関係を使う表現¹¹は評価値が低かった。

3 参照表現の生成手法

主要部後置型である日本語においては、タグ

$$\{G_1, G_2, \dots, G_n, \{x\}\}^{12} \quad (1)$$

に対応する言語表現は、

$$E(G_1) + E(R(G_1, G_2)) + E(G_2) + \dots + E(R(G_n, \{x\})) + E(\{x\}) \quad (2)$$

と表せる。ここで $E(X)$ は X の言語表現、 $R(X, Y)$ は X に対する Y の関係、 $+$ は文字列結合を表す。

ただし全体集合は明示的に参照されない場合が多く、2.2節の実験の分析からも、全体集合を明示的に参照し

⁹布置によっては代表される表現の種類が少なく、選択表現は10個に満たない。また表現は主にタグの異なるものから選んだ。

¹⁰評価値が高くても、正解率が7割未満のものは不適切な表現とした。

¹¹「手前の2つの赤い玉の後ろに並ぶ3つの玉～」等

¹² $G_1 \sim G_n$ は集合、 $\{x\}$ は対象物体のみを含む集合である。

ない表現の評価値の方が高いことが分かった。よって以下で提案する参照表現の生成手法では全体集合にあたる表現は生成しない。

提案手法は次の手順で参照表現を生成する。

step1 物体の位置情報から近接性を基に集合を形成

step2 それらの集合の組み合わせからタグを生成

step3 生成されたそれぞれのタグに表現とスコアを付与

以下に具体的なアルゴリズム（図3～図4）とその説明を記す。

main (図3) 上記の step1 に対応するのは図3の2行目で、関数 `makeGroupList` を呼び出す。この関数は、Thórisson [6] の手法を用いている。この手法は物体同士の近接性から集合を生成するものである¹³。この関数は、生成した集合のリストを返す。ただし対象物体を含まない集合は除き、集合を大きい順に並べ、最後に対象物体のみの集合 $\{x\}$ を追加してある。

例として図2の布置を入力すると、集合のリスト

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

が生成され、上記の処理により、生成されるリスト GL は

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

となる。

そして step2,3 に対応して4行目でタグの生成、表現とスコアの付与を行なう関数 `makeExpressionList` を呼び出す。

makeExpressionList (図4) 4行目、関数 `makeTag` でタグを生成し、5行目、6行目の関数 `setExpression`, `setScore` でそれぞれ表現の設定、スコア付けを行なう。与えられた集合のリストの大きさが n の場合、タグは 2^{n-2} だけ生成される¹⁴。

makeTag 与えられた集合のリスト GL から集合を選択し、それらをリスト (TL とする) としてタグを生成する。タグには全体集合が必要となるため、まず GL から第1要素 (全体集合) を取り出し TL に追加する、次に、 GL の要素数を n とすると、第2要素から第 $n-1$ 要素までの集合を対象にその組み合わせを選択し TL に追加する¹⁵。式 (1) のように、タグの最後の要素は対象物体のみの集合 $\{x\}$ となるので、 GL の第 n 要素 ($\{x\}$) を TL に追加してタグを形成する。

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

¹³この論文のアルゴリズムの表記には誤りがある。正しいコードは Thórisson 氏から入手できる。

¹⁴参照表現の生成に用いるタグには、全体集合と対象物体のみの集合 $\{x\}$ が必要不可欠であるため、タグは残りの $n-2$ 個の集合の組み合わせとなる。

¹⁵図2に対する集合のリスト GL を例とすると、 $\{a, b, c, d, x\}$, $\{a, b, x\}$ がこれにあたり、その組み合わせは 2^2 通り考えられる。

setExpression 与えられたタグに表現を与える作業を行う。これはタグの式 (1) に表現の式 (2) を与える作業に対応する。タグに対応する表現もリスト形式で生成する (EL とする)。まず対象物体 x の表現 $E(\{x\})$ ¹⁶ を EL に追加する。次に $\{x\}$ とそれを含む次に大きな集合 G_n から、それらの位置関係 $R(G_n, \{x\})$ を計算し、その表現 $E(R(G_n, \{x\}))$ ¹⁷ を EL に追加する。さらに、集合 G_n の表現 $E(G_n)$ ¹⁸ を EL に追加する。同様の方法で $E(R(G_{n-1}, G_n)), E(G_{n-1}), E(R(G_{n-2}, G_{n-1})), \dots, E(R(G_2, G_1)), E(G_1)$ を EL に追加する。全体集合は明示的に表現しないので $E(G_1)$ はここで除き、 EL の要素を全て逆に並べ替えて表現 EL を生成する。

例： $\{\text{左奥の、3つの玉のうち、一番右にある、玉}\}$

setScore 与えられたタグのスコアを計算する。ここで集合の規模を表す値として、その集合内の物体の重心となる座標から各物体への距離の平均値を用いる¹⁹。

スコアは、集合 G_i から集合 G_{i+1} へ絞り込む場合、 G_i と G_{i+1} の規模の大きさの比率から算出する。この算出式を決定するために、2.2節で集めたデータの中から包含関係のみを用いた絞り込みの例²⁰を抽出し、その表現の評価値と絞り込みの各集合の大きさの比率を分析した。この関係を図5に示す。三角の点と四角の点は、それぞれ集合から集合への絞り込みの例と集合から対象物体への絞り込みの例である。またこれらの点より得られた2次の近似曲線を示した。この曲線より、集合への絞り込みと対象物体への絞り込みは、それぞれ集合の大きさが $1/4$ と $1/3$ になるように絞り込む表現が評価が高いことがわかる。

実装プログラムには、このグラフから得られた2次の近似式をスコア付けの式として用いた。このスコアを $score(G_1, G_2)$ とすると、最終的なスコアは

$$\frac{1}{n} \left[\sum_{i=1}^{n-1} \{score(G_i, G_{i+1})\} + score(G_n, \{x\}) \right] \quad (3)$$

となる。

4 考察

実装プログラムに、図2の布置を与え、玉 x に対する参照表現を生成させたところ、スコアの高い順に下に示す出力が得られた。

1. 左奥の3つの玉のうち一番右の玉
2. 左の5つの玉のうち奥の3つの玉の一番右の玉

¹⁶今回の実験の場合「玉」となる

¹⁷ここでは2節の実験を参考にして、位置関係 $R(G_n, \{x\})$ に当てはまる表現を1つ選ぶ。例：「一番右にある」

¹⁸本論文では包含関係のみを扱うので、これに表現「のうち」を加える。例：「3つの玉のうち」

¹⁹対象物体のみの集合 $\{x\}$ は重心からの距離が存在しないので定数値を用いた。

²⁰集合から集合への絞り込みの例が26個、集合から対象物体への絞り込みの例が67個存在した

```

0i   #物体
X    #指し示すべき物体
OL   #物体のリスト
GL   #物体の集合のリスト
TL   #3つ組<T,E,S>のリスト

01:OL = { 01, 02, ... ,0n }
02:main( OL, X )
03:  GL = makeGroupList( OL, X )
04:  TL = makeExpressionList( GL )
05:return TL

```

図 3: main

```

T    #タグ：物体の集合のリスト
EL   #表現のリスト
S    #スコア
TL   #3つ組<T,E,S>のリスト
N(a) #aの要素の数

01:makeExpressionList( GL )
02:  TL = {}
03:  for i=1 to 2^( N(GL)-2 )
04:    T = makeTag( GL, i )
05:    EL = setExpression( T )
06:    S = setScore( T )
07:    TL = append( TL, <T,EL,S> )
        #リストの最後に要素を追加
08:  end for
09:return TL

```

図 4: makeExpressionList

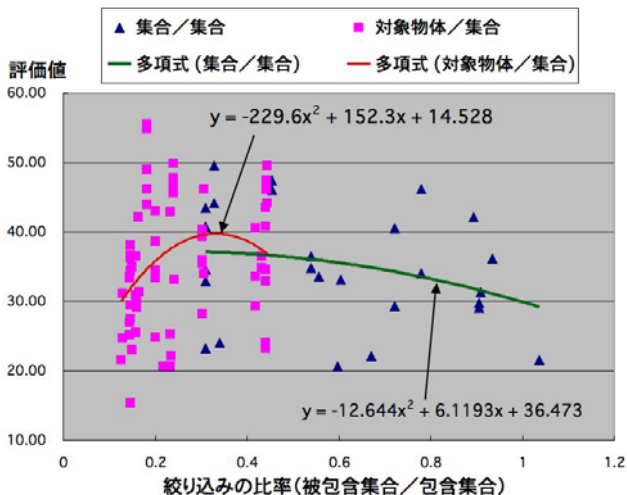


図 5: 絞り込みの比率と評価値との関連

3. 左の5つの玉のうち奥から2番目の玉
4. 右から4番目の玉

これらの表現は図2の x を特定できている。また2節の実験に用いた12種の布置に対する出力は、全て対象物体を特定することができた。しかし提案手法では、生成した表現が対象物体を特定できているかどうか検証していないので、それを定性的に確かめる手法が必要となる。また、今回の実験に用いた布置以外の状況で適切な参照表現が生成されるかは、今後定量的に評価する必要がある。

また、本論文で利用した集合の認識アルゴリズム [6] では「列」を認識できないため、布置によっては思わしくない結果が見られた。

その他、今後の課題として次の2つが挙げられる。

位置関係以外の関係の利用 今回は集合内の関係として位置関係のみを用いたが、「2番目に大きな玉」等の様に大きさや色の程度から集合内の物体を特定する表現も生成できるようにしたい。

従来の参照表現生成手法との融合 本論文で提案した手法は包含関係にある集合間の関係のみを用いて参照表現を生成する。これによって従来の手法では表現できなかった状況も扱うことが可能になった。今後は、提案手法を包含関係にない集合間の関係も扱えるように拡張し、従来手法のように個別の物体間の参照も扱えるようにしたい。

参考文献

- [1] Robert Dale. Generating referring expressions: Constructing descriptions in a domain of objects and processes, 1992. MIT Press, Cambridge.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] Emiel Krahmer, Sebastiaan van Erk, and André Verleg. Graph-based generation of referring expressions. *Computational Linguistics*, Vol. 29, No. 1, pp. 53–72, 2003.
- [6] 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.
- [7] I. van der Sluis and E. Krahmer. Generating referring expressions in a multimodal context: An empirically oriented approach, 2000. Presented at the CLIN meeting 2000, Tilburg.