

非言語情報を用いた協調作業対話における照応解析*

小林 俊平[†] 飯田 龍[†] 徳永 健伸[†] 船越 孝太郎[‡] 中野 幹生[‡]

[†] 東京工業大学 大学院情報理工学研究科

{skobayashi,ryu-i,take}@cl.cs.titech.ac.jp

[‡] (株) ホンダ・リサーチ・インスティテュート・ジャパン (HRI-JP)

{funakoshi,nakano}@jp.honda-ri.com

1 はじめに

頑健な言語理解・生成システムを作る上で、対象を指し示すために使われる種々の参照表現の照応解析は必須のタスクであり、これまでに照応解析について数々の研究がなされている [1]。例えば、構文解析木を用いた Hobbs アルゴリズム [2] や、話題の中心にあるエンティティを考慮する Centering アルゴリズム [3]、統計的機械学習に基づく手法 [4] などがある。しかし、その多くは書き言葉の文章中の照応解析に着目して研究を進めており、対話を対象にした研究は少ない。また、近年の書き言葉を対象とした照応解析に関する研究では、文章内に先行詞が出現する文脈照応が主な研究対象であり、文章外の要素を参照する場合にはその照応詞を解析の対象から除外する傾向がある。しかし、正確に照応現象を把握するためには、談話の流れと現実 (もしくは仮想) 世界の状況に対応付け、文章外照応の場合に現実世界のどの実体を指しているかを理解した上で次の文脈を理解するといった段階的な処理が必要となると考えられる。より現実に即した照応解析を行うためには、文脈情報だけではなく、指差しや話者の行動履歴といった複数のモデルを同時に考慮して解析をおこなうことが必要である。

本研究では、現実世界と発話をリンクさせた状況の一例として、2 人の人間が協調的にパズルを解く際の対話中に出現する照応現象に着目し、パズルのピースの状況やマウスの移動といった非言語情報がどのように照応現象に影響するかを調査した。具体的には、非言語情報を素性として学習を行い、これらの素性が照応解析の精度にどのように寄与するかを定量的に評価した。

本稿では、まず 2 節で対話コーパス作成の実験環境とデータの詳細を説明し、3 節で機械学習に基づく照応解析手法の概要を示す。次に 4 節で従来の照応解析の手法に非言語情報を素性として加えた場合の解析精度の変化について述べ、5 節でまとめる。

2 協調作業対話コーパス

2.1 コーパス作成の実験環境

今回使用したコーパスは、2 人 1 組のペアにタングラムパズルを作業課題として与え、協力して解いた際の様子を記録したものである [5]。ここでタングラムパズルとは、7 つのピースを用いて問題として提示された形を作るパズルである。

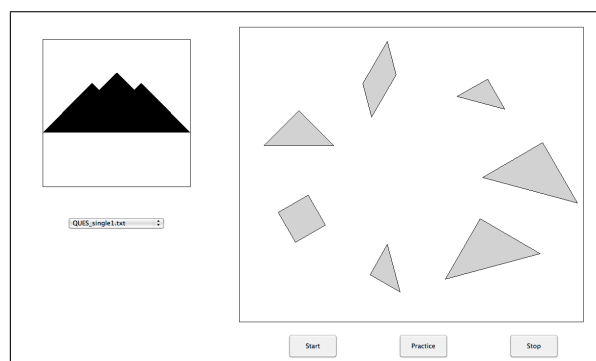


図 1: タングラムシミュレータの画面

被験者の 2 人にはそれぞれ「指示者」と「作業者」という異なる役割を与えた。指示者は提示された目標図形 (図 1 左側) を見ながらその形を形成するピースの配置を考え、ピースの操作を作業者に口頭で指示する。一方、作業者はその指示に従って実際にマウスでピースを操作する (図 1 右側)。目標図形は指示者のみに提示され、操作マウスは作業者のみが扱うことができる。このような非対称な役割を与えることにより、言語表現でのやり取りを促し、より多くの参照表現を収集できる。この方法で収集した計 40 対話を本研究で使用する。

2.2 コーパスの規模

収集したデータの規模は、総対話時間が 417 分 (25,020 秒) であり、1 対話平均は約 10.5 分である。また、対話中に発話された参照表現の総数は 2,363 表現である。つまり平均約 10.6 秒に 1 回の頻度で参照表現が発話されている。

このコーパスにはピースを指示対象とする参照表現にタグが付与されており、その指示対象の個数に応じて 3 種類に分類することができる。1 つ目は特定の 1 つ

* Anaphora resolution in collaborative task dialogue using extralinguistic information

Syunpei Kobayashi[†], Ryu Iida[†], Takenobu Tokunaga[†], Koutaro Funakoshi[‡], and Mikio Nakano[‡]

[†] Tokyo Institute of Technology

[‡] Honda Research Institute Japan Co., Ltd.

表 1: 参照表現の特徴と使用例

特徴	使用数	使用例
1) 指示詞		
連体詞	316	あの 右側の三角形
代名詞	1,018	これ
2) 属性		
サイズ	343	ちっちゃい 三角形
形状	817	大きい 三角形
向き	7	下向いてる 三角形
3) 空間関係		
投影的表現	187	左の 大きい三角形
位相的表現	6	大きい 離れてる やつ
重なり	2	その 下にある 三角形
4) 動作への言及	113	右上に どけた 三角形
5) その他		
残り	22	残りの 大きい三角形
同類	16	それと 同じ形 の

のピースを指す表現であり、2つ目は複数(2個以上)のピースを指している表現である。3つ目は、指示対象が明確に1つのピースに定まらない表現であり、話し手がオブジェクトのタイプ(例. 大きい三角形)のみ示し、インスタンスの選択(例. 2つの大きい三角形のうちどちらを選ぶか)を聞き手に任せている場合がこれに相当する。本研究では、指示対象の数が1つの表現のみを対象に照応解析の問題を扱う。指示対象が1つである表現は、40対話で2,048表現あり、全体の約86.7%である。

2.3 参照表現の分類

2.2で説明した指示対象が1つである参照表現は、以下のような統語的、意味的特徴により細分類できる。

- 1) 指示詞、2) オブジェクトの属性、3) 空間関係、4) オブジェクトに対する動作への言及、5) その他。
- それぞれの特徴の細かい分類と使用数を、使用例とともに表1に示す(1つの表現が複数の特徴を含むため、表現の数は単なる足し合わせにはならない)。

3 照応解析手法の概要

この節では、機械学習を用いて参照表現の指示対象を自動的に同定するモデルの概要と、用いた素性について説明する。

3.1 指示対象の同定

2.2で説明した指示対象が1つである参照表現2,048表現を対象とし、参照表現と各ピースが照応関係となるか否かの分類器を作成する。具体的な同定方法は、以下の通りである。

1. 各参照表現について、その表現の発話が開始された瞬間までの7つのピースの状況を、4節で後述する素性を用いて表現し、7つの特徴ベクトルを作成する。
2. Ranking SVM¹[6]を用いて作成したランカーを利用して、その7つのピースをランク付けし、1位になったピースをその表現の指示対象とする。

上記の手順で各参照表現につき1つのピースを指示対象として対応付けする。

¹http://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html

表 2: 対話中のオンマウス情報の用いられ方

	代名詞	代名詞以外	総数
参照表現総数	548	693	1,241
発話開始時刻に何らかの ピースにオンマウス	491 (89.6%)	504 (72.7%)	995 (80.2%)
発話開始時刻に指示対象の ピースにオンマウス	452 (82.5%)	155 (22.4%)	607 (48.9%)

3.2 代名詞とそれ以外を区別した学習

本研究では、ランカーを作成する際に参照表現を代名詞とそれ以外(以後「代名詞以外の表現」と呼ぶ)をあらかじめ区別して利用する。これは、表1に示した参照表現の特徴のうち、代名詞という特徴は他の参照表現と比較して直示や先行詞との時間的な近さの影響を受けやすいため、学習される特徴が異なるためである。Philippら[7]が本研究で利用するコーパスに対して分析を行ったところ、代名詞は特に直示と共に利用される傾向があることを報告している。また、コーパスの規模が十分に大きければ、素性の組み合わせを考慮できるカーネルを利用することで一度に学習できると考えられるが、評価実験で利用する参照表現の総数は2,048表現であり、そのうち代名詞は1,018表現であるため、コーパスの規模は大きいとはいえない。このため、代名詞と代名詞以外の表現のそれぞれについてランカーを作成し、これにより精度がどのように変化するかを調査する。

3.3 非言語情報

本研究では、学習に利用する素性として、テキスト情報に加え、主に2つの非言語情報を用いる。この項ではその2つの非言語情報の定義について説明する。

3.3.1 オンマウス情報

1つ目の非言語情報は、作業者の動かすマウスがピースの上に乗っているというものであり、これを以後「オンマウス情報」と呼ぶ。これは、人間同士の実環境における直示(指さしなど)の情報に近いものとみなすことができる。

今回のタングラムパズルのタスクにおいて、オンマウス情報は有用な情報になっている。表2に示すように、後に訓練データとして用いる24対話中の参照表現1,241表現において、参照表現の発話開始時刻に指示対象のピースにオンマウスしていたものは607表現(約48.9%)あった。特に代名詞においてこの傾向は顕著で、548表現中452表現(約82.5%)が発話開始時刻に指示対象のピース上にカーソルが位置していた。この結果より、人間が代名詞を発話するときは、直示を伴っていることが多いということが分かる。これに対し、代名詞以外の表現に関しては693表現中155表現(22.4%)と、代名詞に比べて極端に割合が低い。これは、代名詞以外の表現は直示を用いず、他の情報(テキスト情報など)を用いて照応解析をしている場合が多いことを示唆している。ただし、作業者はタスク中考え込んだときなどに無作為にマウスを移動させることがあり、その場合にはまったく関係のないピースの上にマウスカーソルが位置している場

合がある．この点で，オンマウス情報は必ずしも照応関係と関連しているとは限らない．

3.3.2 操作情報

2つ目の非言語情報として，作業者が何らかのピースを操作しているという情報を利用する．これを以後「操作情報」と呼ぶ．今回使用したタングラムパズルのシミュレータでは，作業者はピースの移動，ピースの回転，ピースの反転の3つの操作を選択して行うことができるが，以降ではこの3つの操作を区別せず，操作を行っているか否かという粒度でこの情報を扱う．

3.4 素性

学習・分類には，(a) テキスト情報の素性，(b) オンマウス情報の素性，(c) 操作情報の素性の3種類の素性を用いた．それぞれの素性の詳細を表3に示す．

素性 D9 の「照応解消する表現が代名詞の時，その直前にピースが代名詞以外の表現で参照されているか」という素性を使用する意図としては，「代名詞以外の表現 → 代名詞」の順に発話された場合，その2つの表現は同じ指示対象を指すことが多く，逆に，素性 D10 のように「代名詞 → 代名詞以外の表現」の順に発話された場合，その2つの表現は違う指示対象を指すことが多いと考えたためである．

4 評価実験

前節で説明した照応解析手法を用い，コーパス中に出現する参照表現の指示対象をどの程度自動的に同定することができるかについて評価実験を行った．この節では，設定した実験モデルの説明と，解析結果について説明する．

4.1 実験モデル

3.1 で説明した同定方法を用いて各参照表現に指示対象を対応付けしたものと，コーパスに人手でタグ付けした正解との精度を調べる．収集した40対話のうちの24対話中の参照表現1,241表現(約60.6%)を訓練データ，残りの16対話中の参照表現807表現(約39.4%)をテストデータとした．ただし，今回の実験では Ranking SVM は線形カーネル，パラメタ c は 1.0 に設定して評価を行っている．

また実験を行うに当たって，以下の3つの実験モデルを用意した．

1. 代名詞のモデル: 代名詞のみを学習事例とし，指示表現が代名詞の場合に適用する照応解析モデル
2. 代名詞以外のモデル: 代名詞以外の表現のみを学習事例とし，指示表現が代名詞以外の表現の場合に適用する照応解析モデル
3. 統合モデル: 参照表現全体を訓練事例として学習し，すべての参照表現に対して適用する照応解析モデル

以上3つのモデルに対してそれぞれ実験を行い，その結果の差異について調査した．

表 3: ピースの状況を表す素性

(a) テキスト情報の素性	
D1: yes, no	直前で言及されたピースか
D2: yes, no	以前に言及されてからたった時間が 10 秒未満のピースか
D3: yes, no	以前に言及されてからたった時間が 10 秒以上, 20 秒未満のピースか
D4: yes, no	以前に言及されてからたった時間が 20 秒以上のピースか
D5: yes, no	以前に一度も言及されたことのないピースか
D6*: yes, no, unknown	表現の属性(形, 大きさ)が, ピースの属性と矛盾しないか
D7: yes, no	直前のそのピースを指す表現が, 三格として用いられていたか
D8: yes, no	直前のそのピースを指す表現が, 二格として用いられていたか
D9: yes, no	照応解消する表現が代名詞の時, その直前にピースが代名詞以外の表現で参照されているか
D10: yes, no	照応解消する表現が代名詞以外の表現の時, その直前にピースが代名詞で参照されているか
(b) オンマウス情報の素性	
M1: yes, no	発話開始時刻にオンマウスされているピースか
M2: yes, no	発話開始時刻にオンマウスされていないとき, 直前にオンマウスしていたピースか
M3: yes, no	以前オンマウスされてから経った時間が, 10 秒未満のピースか
M4: yes, no	以前オンマウスされてから経った時間が, 10 秒以上 20 秒未満のピースか
M5: yes, no	以前オンマウスされてから経った時間が, 20 秒以上のピースか
M6: yes, no	以前に一度もオンマウスされていないピースか
(c) 操作情報の素性	
A1: yes, no	発話開始時刻に操作されているピースか
A2: yes, no	発話開始時刻に操作されていないとき, 直前に操作していたピースか
A3: yes, no	以前操作されてから経った時間が, 10 秒未満のピースか
A4: yes, no	以前操作されてから経った時間が, 10 秒以上 20 秒未満のピースか
A5: yes, no	以前操作されてから経った時間が, 20 秒以上のピースか
A6: yes, no	以前に一度も操作されていないピースか

*D6 の yes は属性が矛盾していない場合, no は矛盾している場合, unknown は表現に形と大きさの属性が共になく(例. それ)判断できない場合をそれぞれを表す．

4.2 実験結果

まず，テキスト情報の素性(a)のみを用いた場合(ベースライン)の結果を，表4に示す．次に，ベースラインに2つの非言語情報の素性((b), (c))を加えた場合の結果を表5に示す．さらに表6には，代名詞のモデルと代名詞以外のモデルの各素性の重みをそれぞれ示す．

これらの結果より，まずベースラインに非言語情報を加えることにより，精度が向上することがわかる．特に代名詞に関しては顕著であり，25%以上精度が向上している．また，代名詞以外の表現についても精度が向上しているが，代名詞の場合と比べ精度向上の割合は小さい．これは代名詞以外の表現では非言語情報ではなく，ピースの属性(大きさ, 形など)を用いて照応解析を行っている場合が多いためだと考えられる．評価事例を人手分析したところ，代名詞を用いても良いような場面であえて冗長な表現として代名詞以外の表現を用いてるような場合(例. [オンマウスしているピースに対して] その三角形を...)に精度の改善が見られていた．

また，代名詞と代名詞以外の表現と一緒に学習した統合モデルと個別に学習したモデルについては，テキスト

表 4: テキスト情報のみを用いた場合 (ベースライン)

	総数	正解数	精度
1. 代名詞のモデル	470	290	61.7%
2. 代名詞以外のモデル	337	212	62.9%
1+2*	807	502	62.2%
3. 統合モデル	807	501	62.1%

*「1+2」は、代名詞のモデルと代名詞以外のモデルの結果を足し合わせたものである。

表 5: ベースラインに非言語情報を加えた場合

	総数	正解数	精度
1. 代名詞のモデル	470	411	87.4%
2. 代名詞以外のモデル	337	242	71.8%
1+2	807	653	80.9%
3. 統合モデル	807	604	74.8%

情報のみを利用した場合にはほぼ同様の結果であったが、非言語的情報を加えることにより個別モデルで分類したほうが精度がよくなった。これは、表 6 より、オンマウス情報の素性 (特に M1) など非言語情報の各素性の重みが代名詞と代名詞以外の表現で異なるため、一緒に学習することでこれら特徴を捉えることができなくなるためだと考えられる。

最後に、今回用意したコーパスでは表 6 に示したいくつかの素性がデータスパースネスを起こし、学習することができなかった (代名詞のモデルの A6 など)。これがコーパスの規模が小さいためなのか、もしくは素性の設計に問題があるのかという点については、今後コーパスの規模を拡大するとともにさらに調査を続ける予定である。素性の設計で 1 つ改善が必要と思われる点として、今回は「以前言及されてからの時間」など時間という尺度の素性を用いて学習を行ったが、今回使用したコーパスでは指示者が考え込んだときなどに対話が数十秒間無言になることがあり、この場合時間を素性にするとノイズになってしまう可能性がある。そこで、「時間」ではなく「ディストラクタの数²」とすることで、このようなノイズを取り除くことができるのではないかと考えている。

5 おわりに

本稿では、協調作業対話における参照表現の指示対象の同定に、テキスト情報だけでなく、マウスカーソルによるポインティングや操作履歴といった非言語的な情報を用いて機械学習を行う手法を提案した。提案手法を用いて評価実験を行った結果、テキスト情報だけを利用した場合に比べ照応解析の精度は最大で 25% 以上 (代名詞のモデルの場合) 向上することが分かった。また、代名詞と代名詞以外の表現では利用すべき情報が異なるため、個別に分類するモデルを作成したほうが解析精度が向上することを明らかにした。

今後の課題としては大きく 3 つの方向性を考えている。1 つは今回は 2 人の発話者がタングラムパズルを解く際の対話を対象に照応解析の手法を提案したが、一部の手がかりは領域依存であるため、今後は他の種類の対

²タングラムのタスクでは、指示対象が以前言及されてから現時点までの間に含まれる他のピースへの参照の数のことである。

表 6: 別々に分類した場合の各素性の重み

ランク	代名詞		代名詞以外	
	素性	重み	素性	重み
1.	M1	0.6780	D6	0.6918
2.	M3	0.2749	D2	0.2320
3.	D2	0.2286	M3	0.1639
4.	D1	0.2080	A3	0.1435
5.	A3	0.1290	D1	0.1318
6.	M2	0.1123	M1	0.1239
7.	D9	0.1105	D7	0.0797
8.	D6	0.0365	M4	0.0136
9.	D7	0.0152	D3	0.0126
10.	A2	0.0123	A4	0.0118
11.	A1	0.0109	D9	0*
12.	D10	0*	A1	-0.0003
13.	A6	0	M2	-0.0031
14.	D3	-0.0100	A2	-0.0078
15.	A4	-0.0228	A6	-0.0103
16.	D8	-0.0465	D10	-0.0185
17.	M6	-0.0466	M5	-0.0416
18.	D5	-0.0860	D4	-0.0562
19.	M4	-0.0885	D8	-0.0597
20.	A5	-0.1062	M6	-0.1358
21.	D4	-0.1326	A5	-0.1449
22.	M5	-0.1398	D5	-0.1885

*D9 は条件に「照応解消する表現が代名詞の時」と条件を付けているので、代名詞以外の表現の時には素性の値に yes が 1 つも付かないから当然重みもない。D10 に関しても同様である。

話例も対象に研究を進めることで、領域非依存の手がかりとそれ以外との境界を明らかにしていく必要がある。2 つ目に、コーパスの規模を拡大することにより、今回「代名詞以外の表現」とひとまとめにしたものをさらに細かい特徴を用いて分類して調査、実験を行い、各特徴ごとの表現が発話される際の世界の状況の特異性についても調べていきたい。最後に 3 つ目として、本研究ではピースとマウスカーソルとの位置関係のような非言語情報を利用したが、これに加えユーザの視線のような他の非言語情報も解析に役立つと考えられる [8]。タングラムパズルの対話についてはすでに視線を含めたデータの収集を開始しており、今後これらの情報も含めて照応現象の分析を行う予定である。

参考文献

- [1] Ruslan Mitkov. *Anaphora Resolution*. Longman, 2002.
- [2] Jerry Hobbs. Resolving pronoun references. In *Lingua*, pp. 44:311–338, 1978.
- [3] Susan E. Brennan, Marilyn W. Friedman, and Carl J. Pollard. A centering approach to pronouns. In *Proceedings of the 25th annual meeting on Association for Computational Linguistics*, pp. 155–162, 1987.
- [4] G. Doddington, A. Mitchell, M. Przybocki, L. Ramshaw, S. Strassel, and R. Weischedel. An expectation maximization approach to pronoun resolution. In *Proceedings of the Ninth Conference on Natural Language Learning (CoNLL-2005)*, pp. 88–95, 2005.
- [5] 安原正晃, Philipp Spanger, 飯田龍, 徳永健伸. 協調作業対話で用いられる参照表現の収集と分析. 言語処理学会第 15 回年次学会, pp. 622–625, 2009.
- [6] Thorsten Joachims. Training linear SVMs in linear time. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006.
- [7] Philipp Spanger, Yasuhara Masaaki, Iida Ryu, and Tokunaga Takenobu. Using extra linguistic information for generating demonstrative pronouns in a situated collaboration task. In *Proceedings of Workshop on Production of Referring Expressions: Bridging the gap between computational and empirical approaches to reference*, 2009.
- [8] Zahar Prasov and Joyce Y. Chai. What's in a gaze?: the role of eye-gaze in reference resolution in multimodal conversational interfaces. In *Proceedings of the 13th international conference on Intelligent user interfaces: International Conference on Intelligent User Interfaces*, pp. 20–29, 2008.