

ポスター会話における指示表現の分析 —参照先との自動対応付けに向けて—

吉田 仙^{†*} 高梨 克也[‡] 河原 達也[‡] 永田 昌明[†]

[†] NTT コミュニケーション科学基礎研究所

[‡] 京都大学学術情報メディアセンター

1 はじめに

会話において現実の場にある実体や会話中で言及された事柄を参照するために「これ」「その本」といった指示表現が用いられる。指示表現は、それが何を参照しているのかが話者の動作や会話の文脈などから明らかかな場合に用いることができ、指示表現を用いることにより発話を簡潔にすることができる。

一方、会話から指示表現を含む文だけを取り出した場合には、それが何を参照しているのかわからなくなってしまう。例えば話者が机を指差しながら「これは木でできている」と言ったとして、その会話の書き起こしからその文を抜き出して読んでも何について話をしているのか理解することはできず、その様子を撮影した映像と照らし合わせるなどを行ってはじめて理解が可能になる。このとき、「これ」という指示表現は映像中にある机のイメージと結びつけられる。

また、指示表現の参照対象がその会話の中で言及されている事柄であれば、指示表現の出現箇所から発話内容を遡ってその対象への言及を探し当てることにより指示表現を含む文の理解が可能になる。このような処理は照応解析の一種であり、書き言葉では統計的機械学習を用いて参照先を自動的に同定する研究が行われている [1]。

このように会話の文中に含まれる指示表現をそれが参照している事物と結びつけることができれば、例えば会議の音声認識結果を提示して議事内容のブラウジングを可能にするツール [4] において、指示表現の出現箇所からその参照先を表す映像や音声認識結果テキストの該当部分にリンクを張るなどの機能を実現することが期待できる。このような理由から、我々は会話における指示表現とその参照先との対応付けを自動的に行う技術の実現をめざした研究を行っている。

本稿では、会話における指示表現の出現傾向を実際のデータをもとに調査した結果を報告する。データとして、図 1 のような学会のポスター発表を模した会話

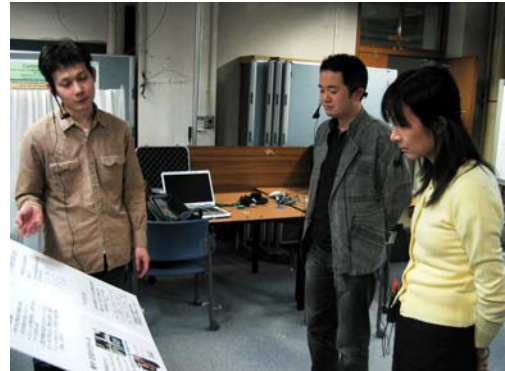


図 1: ポスター会話

[2, 6] の書き起こしを用い、独話や書き言葉のテキストと比較することで会話における指示表現の特徴を明らかにする。またポスター会話の書き起こしと非言語行動データとの関連を調べ、指示表現の参照先自動対応付けに対してどのような情報を活用できるのか検討する。

2 ポスター会話における指示表現

会話における指示表現の参照先は、第 1 章で述べたように、発話の状況下における実体である場合と、会話中で言及された事柄である場合がある。照応解析の観点では、照応詞である指示表現が会話中で言及された事柄を参照している場合を文脈照応、参照先が文中には無い場合を外界照応と呼ぶ。現実の場にある実体を参照している場合でも、先行文脈においてそれを言及しているのであれば文脈照応である。また、照応詞が現実の場にある実体を指し示すことは直示 (deixis) と呼ばれる。したがって指示表現は表 1 に示す 4 種類に分類できる。

照応解析の従来研究で多く見られる先行詞の素性を用いた照応解析は、これらのうち表の左側の部分を対象としていると言える。一方、指示表現とその参照先との対応付けという意味では表の右上もターゲットとなり、発話内容以外の映像などの情報を用いれば解析

*yoshida@cslab.kecl.ntt.co.jp

表 1: 指示表現の参照タイプによる文類

	先行詞あり	なし
場にある実体を指している	文脈照応、直示	外界照応、直示
指していない	文脈照応、非直示	(外界照応、非直示)

できる可能性がある。また左上の範囲については発話内容に対する解析に加えて映像等の情報も利用することで解析の精度を高められる可能性がある。

ところでポスター会話には、それが行われる場にポスターという情報媒体が存在し、その内容を利用して会話が進められるという特徴がある [6]。したがってポスター会話では、表の上半分に該当する参照表現が、情報媒体の無いスタイルの会話・独話やあるいは書き言葉に比べて多く出現することが予測される。

3 ポスター会話データの分析

3.1 異スタイルのテキストとの比較

ポスター会話における指示表現の出現傾向をスタイルの異なるテキストと比較するため、以下の3種類のコーパスに対して調査を行った。

billboard ポスター会話 [2, 6] の書き起こし。会話は説明者1名、聞き手役2名の計3名で行われている。説明者と聞き手役では発話のスタイルに違いがあることが予想されるため両者を分けて調査する。

CSJ 日本語話し言葉コーパス (CSJ) [9] の書き起こしに対して照応のタグを付与したもの [5]。CSJの独話には学会講演と模擬講演がある。学会講演では発表用スライドが用いられるため、独話とはいえずポスター会話におけるポスターのような情報媒体が存在するといえる。一方、模擬講演ではスライドは用いられていない。このため学会講演と模擬講演を分けて調査する。

NAIST NAIST テキストコーパスに指示表現のタグを付与したもの [1]。新聞記事、すなわち書き言葉のコーパスである。

上記3種類のコーパスの各種属性を表2に示す。

これらのコーパスに対し、まず指示表現の出現頻度を調査した。3種類のコーパスのうちCSJとNAISTのコーパスで共通してタグ付けされている指示代名詞「これ」「それ」「あれ」および指示連体詞「この」「その」「あの」について出現頻度を調査した結果を表3に示す。

表3より、ポスターや発表用スライドという情報媒体が存在する billboard および CSJ 学会講演では、情報媒体が存在しない CSJ 模擬講演や NAIST に比べて指示表現の出現頻度が多くなっており、特に「これ」「この」というコ系の指示表現においてその傾向が顕著である。このことから billboard や CSJ 学会講演ではポスターや発表用スライドを指す直示が多数用いられていることが推測される。

また、「それ」という指示表現が、特に billboard 聞き手役において多く見られる。情報のなわばり理論 [8] の観点からすると、ポスター会話における会話の内容は基本的に説明者の研究に関することであるため、聞き手役にとっては相手側のなわばりの中にあるものである。このことからソ系の指示代名詞である「それ」を多く使っていると推測することができよう。ただし聞き手役の発話は1文あたり文節数を見れば分かるように短いため、1回の発話でどれくらい「それ」を使うかという意味では他のスタイルのテキストと比べて大きな差は無い。「これ」「この」が説明者に比べ聞き手役で多くなっているのも1発話あたりで考えれば差は無いといえる。

3.2 非言語行動データとの関連性

billboard ポスター会話コーパスでは、会話の書き起こし以外にも様々なモーダルデータが提供されている。この節では非言語モーダルデータと指示表現との関連性について述べる。

はじめに、指差し動作のアノテーションデータと指示表現との関連性について調べる。このアノテーションデータは、ポスター会話の各参加者の手の動きをビデオで観察し、ポスターを指差している状態¹、または指差していない状態のいずれかにラベル付けしている。

ポスター会話に現れる指示表現のうち「これ」「この」といったコ系のものは、第3.1節での調査で情報媒体の有無による頻度の差が見られたことから、直示としてポスターを指差しながら発せられることが多いことが予想される。そこで、書き起こしにおけるコ系の指示表現と指差しアノテーションデータにおける指差し状態の共起をカウントする。ここで、指差しアノテーションデータでは各状態の開始/終了時刻が記録され

¹実際のデータでは4枚あるポスターのそれぞれを指差している状態、およびポスターが並んでいるあたり全体を指差している状態に区分されているが、ここではそれらをまとめてポスターを指差している状態とする。また、それらの状態への遷移のために手が動いている状態もそれらの状態に含める。

表 2: コーパスの各種属性

	対象範囲	文区切り	指示表現の認定
billboard	節単位や指差し動作などのタグを付与した4会話	節単位タグ(あいづち部分は除外)	自動(品詞、表記)(人手で認定されたフィラーを除く)
CSJ	照応タグを付与した学会6講演、模擬6講演	節単位タグ	人手
NAIST	報道記事 883、社説記事 580 の計 1,463 記事	文タグ	人手

表 3: 指示表現出現頻度のコーパス間比較

		文節数		1,000 文節あたり指示表現数						
		総数	1 文あたり	これ	それ	あれ	この	その	あの	計
billboard	説明者	7,503	9.62	7.46	10.53	0.53	13.99	10.13	0.00	42.65
	聞き手役	1,569	3.86	13.38	16.57	0.64	20.40	7.65	0.00	58.64
CSJ	学会講演	5,822	10.92	12.37	4.81	0.00	18.55	7.90	0.00	43.63
	模擬講演	5,359	12.26	3.36	7.09	0.00	4.29	11.76	0.00	26.50
NAIST		285,598	9.58	3.82	2.84	0.04	6.32	6.18	0.13	19.34

ているが、書き起こしでは IPU²の開始/終了時刻しか時刻情報が得られないため、指示表現を1個以上含む IPU と指差し状態の共起をカウントする。IPU 区間に指差し区間が一部でも重なれば共起とみなす。

その結果、表 4 に示すように、コ系指示表現を含む IPU 区間では含まない区間に比べポスター指差し動作と共起する割合が顕著に増加していた。

表 4: ポスター指差し状態との共起割合 (%)

コ系指示表現を含む IPU	58.3
コ系指示表現を含まない IPU	31.9

次に、視線のアノテーションデータと指示表現との関連性について調べる。このアノテーションデータでは各参加者の注視先を個々の参加者、聞き手全体、ポスター、その他のいずれかにラベル付けしている³。そこでコ系の指示表現を1個以上含む IPU とポスター注視の共起をカウントする。

この結果、表 5 に示すように、コ系指示表現を含む IPU 区間では含まない IPU 区間に比べポスター注視と共起する割合が増加していた。

表 5: ポスター注視との共起割合 (%)

コ系指示表現を含む IPU	84.8
コ系指示表現を含まない IPU	62.9

以上の結果から、コ系の指示表現は直示として用い

られることが多いことが非言語データとの関連という点からも示唆される。

3.3 「この」を受ける名詞の調査

情報媒体の有無により出現頻度に差が見られたコ系指示表現において、どのような対象を直示しているのかを知る手がかりとして、「この」+名詞句という構文における名詞句部分の意味カテゴリを調査した結果を表 6 に示す。参考のために「この」を受ける名詞句以外も含めたテキスト内の全名詞の分布も併せて示す。なお、意味カテゴリおよび係り受け情報の付与には JTAG・Jdep [7] を用いた。

これによると、情報媒体が無い場合と有る場合、さらには指差しと共起している場合で、やや分布のばらつきは有るもののさほど大きな傾向の変化は見られなかった。このことはどのような事物でも直示における指示対象になりうることを示していると考えられることができるが、今後さらにデータ量を増やし、また名詞の分類方法を工夫したりポスター上に書かれた文字との関連を調べる等を行い、指示対象特有の傾向を探る必要がある。

4 まとめ

本稿では、指示表現の参照先の自動対応付け技術の実現に向け、ポスター会話の書き起こしを対象として

²転記基本単位。原則として 200ms 以上のポーズで区切られた範囲。

³各状態への遷移途中の状態はここでは各状態に含める。

表 6: 「この」を受ける名詞句の主辞の意味カテゴリ分布 (レベル 2 まで) (%)

		名詞	具体	主体	場所	具体物	抽象	抽象物	事	抽象的關係
全名詞	情報媒体なし	1.0	0.0	18.6	9.7	5.0	0.0	10.7	29.4	25.6
	情報媒体あり	4.6	0.0	17.9	5.7	13.3	0.0	6.5	24.3	27.7
この+X	情報媒体なし	0.1	0.0	6.9	9.3	4.6	0.0	19.4	24.5	35.3
	情報媒体あり	0.4	0.0	11.3	6.5	4.8	0.0	19.1	10.9	47.0
	指差し共起	0.0	0.0	17.9	9.5	5.3	0.0	20.0	13.7	33.7

(情報媒体なし: CSJ 模擬講演+ NAIST、情報媒体あり: billboard + CSJ 学会講演、指差し共起: billboard のうち指差しと共起する「この」+X のみ)

実際のデータにおける指示表現をいろいろな側面から調査した。

第 3.1 節で述べたように、ポスターという情報媒体が存在する billboard データではコ系指示表現の頻度が高く、直示が多く用いられていることが推測された。したがってポスター会話での参照先自動対応付けにおいては、第 2 章の表 1 の上半分に対する参照先推定の重要性が書き言葉に比べて増している。第 3.2 節で示したように指差し動作や視線といった情報と直示との関連が示唆されており、自動対応付けの要素技術にこれらをどう活かすか検討が必要である。

また、ポスター会話での参照先の自動対応付け実現に向けては、その正解データを構築する事も極めて重要である。例えば Müller は英語の多人数会話に出現する指示表現 *it, this, that* について機械学習を用いて参照先を自動推定する手法を提案しているが [3]、このような手法を利用するのであればその学習データや評価用データが不可欠である。また、第 3.3 節で「この」+名詞句の名詞句部分の意味カテゴリを調べたが、正解データを構築すればさらに綿密な分析を行うことができる。例えば文脈照応と外界照応の分布や指定指示 (直接照応) と代行指示 (間接照応) [1] の分布に関する調査、あるいはポスター上に書かれた文字とそれを直示参照する発話での表現の関連性の調査などが望まれる。

以上のように、本研究には今後取り組むべき様々な課題がある。

謝辞

NAIST テキストコーパスへの指示表現タグ付与データを提供いただき、また貴重なコメントを頂きました東京工業大学の飯田龍助教、奈良先端科学技術大学院大学の井之上直也氏に感謝します。

参考文献

- [1] N. Inoue, R. Iida, K. Inui, and Y. Matsumoto. Resolving direct and indirect anaphora for Japanese definite noun phrases. *Journal of Natural Language Processing*, 17(1), 2010.
- [2] T. Kawahara, H. Setoguchi, K. Takanashi, K. Ishizuka, and S. Araki. Multi-Modal Recording, Analysis and Indexing of Poster Sessions. In *Proceedings of Interspeech*, pp. 1622–1625, 2008.
- [3] C. Müller. Resolving *it, this, and that* in unrestricted multi-party dialog. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pp. 816–823, 2007.
- [4] S. Tucker and S. Whittaker. Accessing multimodal meeting data: Systems, problems and possibilities. In *Machine learning for multimodal interaction: first international workshop, MLMI 2004 revised selected papers*, Vol. 3361 of *Lecture Notes in Computer Science*, pp. 1–11. 2005.
- [5] S. Yoshida and M. Nagata. Utilizing features of verbs in statistical zero pronoun resolution for Japanese speech. In *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation*, Vol. 2, pp. 867–874. 2009.
- [6] 瀬戸口, 高梨, 河原. 多数のセンサを用いたポスター会話の収録とその分析. *情報処理学会研究報告*, 2007(75):31–36, 2007.
- [7] 今村, 齋藤, 浅野. テキストからの知識抽出の基盤となる日本語基本解析技術. *NTT 技術ジャーナル*, 20(6):20–23, 2008.
- [8] 神尾. 情報のなわ張り理論—言語の機能的分析. 大修館書店, 1990.
- [9] 日本語話し言葉コーパス. <http://www.kokken.go.jp/katsudo/seika/corpus/>.