

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 3 月 31 日現在

機関番号：12608

研究種目：基盤研究(B)

研究期間：2009～2011

課題番号：21300049

研究課題名（和文） 視線情報を利用した参照表現の理解・生成モデル

研究課題名（英文） Understanding and generation of referring expressions using gaze

研究代表者

徳永 健伸（TOKUNAGA TAKENOBU）

東京工業大学・大学院情報理工学研究科・教授

研究者番号：20197875

研究成果の概要（和文）：人間が対象物体を指示する時に用いる参照表現を計算機で扱うモデルを構築するための基礎データとして、日本語の参照表現コーパスを作成した。2名で幾何パズルを協調して解く課題を与え、その時の対話音声、パズル・ピースの位置、パズル・ピースに対する操作、および対話者の視線などの情報をすべて時間同期して記録した。対話音声を書き起し、その中から参照表現を抽出し、指示対象と表現中の性質をアノテーションし、合計 131 対話からなるマルチモーダル・コーパスを作成した。このコーパスを基礎データとし、視線や行動などの非言語情報を利用した参照表現の理解・生成モデルを機械学習の枠組を用いて実現した。作成したコーパスを用いてこれらのモデルの評価をおこない、その有効性を確認した。

研究成果の概要（英文）：

We constructed a Japanese referring expression corpus, which serves as fundamental data for research on computational model of understanding and generation of referring expressions. In our experimental setting, two dialogue participants collaboratively solve geometrical puzzles. In each dialogue, arrangement of puzzle pieces, operations on puzzle pieces and eye gaze of the participants were recorded in synchronization with dialogue speech. Referring expressions were extracted from transcribed speech texts and annotated with their referents and attributes. The constructed corpus includes 131 dialogues in total. Using this corpus, computational models for understanding and generation of referring expressions were constructed based on a machine learning framework. The models were evaluated by using the constructed corpus to show the effectiveness of the models.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009 年度	9,900,000	2,970,000	12,870,000
2010 年度	2,900,000	879,000	3,770,000
2011 年度	1,700,000	510,000	2,210,000
年度			
年度			
総計	14,500,000	4,359,000	18,850,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：参照表現，自然言語処理，協調作業対話，視線情報，幾何パズル，対話システム，状況依存対話，言語生成・理解

1. 研究開始当初の背景

人間と計算機のインタラクションにおいて言語は一義的なモダリティであり、言語処理の分野では対話処理の研究が盛んにおこなわれてきた。しかしながら、これまでの研究では、言語を閉じた記号システムとして扱う傾向が強く、世界と記号システム(言語)との関係についてはそれほど高い関心が払われてきたとはいえない。この傾向は過去の対話システムの多くが旅行のプランニングや情報検索などのいわゆる情報探索型の対話を対象としていることに顕著に表われている。

最近のコンピュータ・グラフィクス(CG)技術やロボット技術の進展は目覚しく、CG合成されたキャラクタやロボット(エージェント)と言語を通してインタラクションする要求が高まっている。エージェントとのインタラクションに関する研究は、1990年代の末から2000年代初頭にかけて米国MITでおこなわれた一連の研究が先駆的なものである。しかし、これらの研究では、自然な動作の生成や、ジェスチャーなどの非言語的な情報の解釈に主な焦点が当てられており、言語の理解という観点からの研究は十分ではない。参照表現の研究に関しては、認知言語学の分野では1990年代終りから視線情報に着目した対話研究はいくつかおこなわれおり、最近では論文誌の特集号も組まれている。しかし、本研究課題で提案するマルチモーダルな状況での参照表現の問題に計算言語学の観点から取り組む研究は緒に就いたばかりで、いくつかの研究グループが研究を始めたばかりであり、未だ十分な成果が出ていない。マルチモーダル、特に視線情報をともなう参照表現のコーパスの作成や分析も試みられてはいるが、その方法論や分析方法については試行錯誤の段階である。また、少なくとも日本語に関してはこのようなコーパスは皆無である。

2. 研究の目的

我々のこれまでの研究成果および、関連分野の過去の研究成果をふまえて、本研究課題では具体的な目標として以下の2つを設定する。

(1) 日本語の対話における話者の視線情報を含む参照表現コーパスの作成

2名の話者が協力しながら与えられた課題を解く際の発話と話者の視線情報を同時に記録した基礎データを収集する。視線情報が対話の理解、特に参照表現の理解・生成にどのように寄与するかを明らかにするために必要な情報を検討し、収集した基礎データに付与し、コーパスを構築する。広く参照表現研究の分野に貢献するために、このコーパスは研究用に公開する。

(2) 視線情報を利用した参照表現の理解と生成モデルの構築

作成したコーパスの分析に基づき、参照表現の理解と生成のモデルを作成し、プロトタイプ・システムに実装し、モデルの妥当性を評価する。

3. 研究の方法

(1) 参照表現コーパスの構築と分析

以下の2段階に分けて研究を実施した。

1組(2名)の話者が協調して与えられた課題を達成する状況を設定する。各話者の発話および視線情報を記録し、それを書き起し、参照表現に関連する情報を付与したコーパス(以下、「参照表現コーパス」と呼ぶ)を作成した。

参照表現コーパスを分析し、参照表現の理解・生成に有用な情報について検討し、それらの情報をコーパスにフィードバックした。実験計画

課題：被験者に与える課題としてタングラム・パズルを用いた。タングラムは、大中小3種類の大きさの5つの直角三角形と正方形、平行四辺形を組み合わせて様々な形を作るパズルである。タングラムを選んだ理由は、扱うピースの数が限られているため、対話中で用いる参照表現の異り数が莫大な数にならないと予想されるためである。これはプロトタイプ・システムを作成する上でも大規模な言語資源を用意する必要がないという点でも利点となる。認知科学の分野でも、同様な幾何学パズルを課題として用いる研究はいくつか報告されており、標準的な課題であると考えられる。

コーパス・サイズ：参照表現のコーパスとしてはTUNAコーパスが有名である。TUNAコーパスは被験者に静的な物体の配置、その中の特定の物体を指示するための参照表現を被験者に生成させることによって、参照表現を収集している。参照表現コーパスとしては先駆的だが、静的な状況しか扱っていない点、状況を表す図と言語表現しか情報として収集されていない点に限界がある。TUNAコーパスの一部は参照表現生成の共通タスクのデータとして公開されており、公開されているコーパス中には780の参照表現が含まれている。Fosterらもタングラムを使ったコーパスを作成しているが、彼らの主な関心はピースに触れる動作と参照表現(行為)の分析である。コーパス・サイズは、512対話を収集し、1,333の参照表現を得ている。本研究課題では、これらの既存のコーパスを越える規模のものを構築するために、条件をいくつか変化させて、合計で170対話を収集し、最終的には8,859の参照表現を含むコーパスを作成した。

役割設定：参照表現をできるだけ自然に誘発するために、2名の話者に非対称な役割を設定した。一方(指示者)には、タングラムのピースを使って作成する目標図形とピースの配置状況を提示し、もう一方(操作者)には目標図形は提示せず、ピースの状況のみを提示した。また、指示者は直接ピースにさわらず、作業者のみがさわられるものとした。ピースの状況は2名の話者に共有されているので、指示者がパズルを頭の中で解き、言葉で作業者にピースの移動を指示し、作業者がその指示にしたがってピースを動かすことになる。これより指示者の発話中にはピースを指示する参照表現が頻出することが期待できる。

記録情報：タングラムのピースの配置は時刻と共に変化するので、物理的なタングラム・パズルの代わりに、コンピュータ・ディスプレイ上に表示したタングラムをマウス操作によって移動・回転・反転などの基礎的な操作ができるシミュレータを作成し、作業者にはこのシミュレータを使ってピースを操作させた。シミュレータを使うことによって、各時刻におけるピースの配置、マウスの位置、操作が正確に記録できる。また、視線計測装置によって、同時に各時刻におけるディスプレイ上の視線位置を記録した。これにより、シミュレータから得られる座標位置と、視線計測装置から得られる座標位置から、ある時刻で話者がどのピースを見ていたかを自動的に判定することができる。話者の発話に関しては、シミュレータと同期をとりながら、2つの接話マイクで話者の発話を左右のチャンネルにステレオ録音し、これをテキストに書き起した。

さらに、実験条件を変えて複数の種類のコーパスを収集した。具体的には、パズルの種類(タングラム、ポリオミノ、ダブル・タングラム)、パズルを解くためのヒントを途中で与えるかどうか、言語の違い(日本語と英語)、の3要因について条件を変更した。

参照表現の抽出・分類

マルチモーダル・アノテーション・ツールELANを用いて、対話の書き起しテキスト中の参照表現を抽出した。参照表現の定義としては、基本的にピースを指している表現とした。参照表現の抽出と同時にその参照表現が指している指示対象のピースも同時に記録した。抽出した参照表現を分類し、各分類の参照表現を理解・生成するために、有用な情報について検討する。その検討をふまえ、さらに付与すべき必要な情報があれば、それらをコーパスに反映し、コーパスを洗練する。

(2) 参照表現の解析モデルの構築と評価

参照表現の理解と生成の2つの側面から計算モデルを構築し、いずれも前述のコーパスを用いて評価をおこなった。いずれのモデルも機械学習の枠組を利用した。参照表現の理解

モデルでは、参照表現の言語的特徴とその表現が使われた談話履歴および視線や行動などを含む使用状況の特徴を素性として、指示対象を識別するという定式化をおこなった。一方、参照表現の生成モデルでは、コーパスの分析結果から代名詞の使用が特徴的であったため、代名詞をどのような状況で使うかを機械学習の枠組でモデル化した。使用した特徴は解析モデルに準じた。いずれのモデルも構築したコーパスを用いた内生評価によって評価をおこなった。

4. 研究成果

(1) 参照表現コーパスの構築

コーパス作成のためのタングラム・シミュレータを開発した。また、タングラム・シミュレータで記録する対話音声、パズル・ピースに対する操作の情報、各パズル・ピースの位置情報と視線計測装置で記録する対話参加者2名の視線情報を時間同期する手法について検討し、実装した。

合計39組(78名)の被験者を使い、1組あたりに4つ(後述するダブル・タングラム課題では6つ)の問題を与え、合計170対話を収録した。このうち視線情報が70%以上取得できた131対話について音声の書き起しをおこない、参照表現のアノテーションをおこなった。アノテーションする情報としては、参照表現の区間、その指示対象、および参照表現の属性である。属性としては、いくつかの対話をサンプリングし、必要だと考えられる属性をあらかじめ用意したが、アノテーションを進める過程で必要に応じて追加・修正をおこない、最終的には以下の属性集合を得た。

dpr: 指示代名詞
dad: 指示形容詞
dmn: 形式名詞
siz: サイズ
col: 色
typ: タイプ
dir: 方向
prj: 投射型空間関係
tpl: 位相型空間関係
ovl: オーバーラップ
act: ピースに対する操作
cmp: 補集合
sim: 類似性
num: 数
rpr: 修正
err: エラー
nest: 参照表現の入れ子
meta: 比喩表現
nul: 上記にあてはまらない場合

さらに対象領域を拡大するために、2つのパズルを導入し、タングラムと同じ実験環境で対話データの収録をおこなった。ひとつは

タングラムのパズル・ピースを2組使いさらにパズル・ピースに色を付けたパズル課題(ダブル・タングラム)を用意した。また、2組のパズル・ピースを用いるために指示対象のあいまい性が増大するので、従来とは異なる種類の参照表現を収集することができた。

もうひとつのパズル課題として7つのピースからなるポリオミノ・パズルを用意した。タングラムと違い、ポリオミノではピースの形状が三角や四角といった単純な幾何形状ではないために、対話者が形状に関しては独自の命名をする必要がある。したがって、より多様な表現を収集することができた。

また、従来のタングラム・パズルでは、主にパズルを解く役割の指示者が深く考え込んで発話をしなくなる状況を避けるために、5分毎に正しいピースの位置を1つだけ示すことによってヒントを与えていたが、問題解決の段階によって使用される参照表現がどのように変化するかを分析するためには好ましくない。この分析のためにヒントを提示しないタングラム・パズルを課題とする対話の収録もおこなった。

英語については従来から参照表現の研究が盛んにおこなわれているが、言語間の比較を検討するための基礎データとして、英国 Brighton 大学の協力を得て、英語話者についてもタングラム課題について(ただし視線情報は除く)6組24対話の収録をおこない、英語版の参照表現コーパスを作成した。

構築した参照表現コーパスの概要を以下に示す。131対話中でアノテーションをおこなった参照表現の総数は8,859であり、このコーパスは、記録している情報の粒度および精度、その規模を考えると、国内はもとより国際的に見ても他に類をみない豊富な情報を付与したマルチモーダル・コーパスである。

課題	組	収録対話数	有効対話数
タングラム (視線なし)	6	24	24
タングラム	8	36	27
タングラム (ヒントなし)	5	20	8
ダブル・タン グラム	7	28	24
ポリオミノ	7	42	24
タングラム (英語)	6	24	24
合計	39	170	131

(2) 参照表現解析モデルの構築

構築した参照表現コーパスの分析に基づいて参照表現の指示対象を同定するモデルを機械学習の手法を用いて作成した。指示対象の同定問題を与えられた参照表現とそれ

が使われた状況を入力として、その参照表現が指示している対象(パズルピース)を可能性が高い順にランキングする問題として定式化した。ランキングには Ranking Support Vector Machine を用いた。参照表現が使われる状況を記述するために、コーパスの分析に基づいて談話履歴素性(10種類)、行動履歴素性(12種類)、視線素性(14種類)を設定した。タングラム課題の27対話分のデータを用いてモデルを評価した結果、利用する素性の組合せ、参照表現の種類によって、以下のような指示対象の同定精度を得た。

この結果からわかるとおり、代名詞の指示対象同定にはその時点でどのような行動をおこなっているかが重要な手がかりとなり、代名詞以外の参照表現の指示対象同定には視線情報が有力な手がかりとなるとという新しい知見を得ることができた。

(3) 参照表現生成のモデルの構築と評価

構築した参照表現コーパスを利用して参照表現を生成するモデルを構築し、コーパスを利用して内生的な評価をおこなった。対象とするドメインでは代名詞の利用が特徴的であることから、与えられた状況においてあるパズルピースを指示するために代名詞を利用することができるかどうかを判断する分類問題として、代名詞の生成を定式化した。分類器としては Support Vector Machine を用い、入力として、参照表現を使用する状況と指示対象を与え、出力として代名詞を使うべきかどうかの二値判断をする分類器を作成した。状況を記述するための素性として、談話履歴素性(5種類)、行動履歴素性(5種類)、実行中の行動に関する素性(2種類)を設定した。タングラム課題のデータ24対話分を用い、評価実験をおこない、以下のような結果を得た。

素性	再現率	精度	F 値
談話+行動+実行中	69.8	67.4	68.6
行動+実行中	73.5	66.1	69.6
談話+実行中	69.8	67.3	68.5
談話+行動	58.7	57.2	57.9

この結果から参照表現理解の結果と同様、代名詞の生成には現在実行中の行動が大きな影響を持つことがわかった。ただし、これはあくまでも内生的な評価なので、与えられた状況でコーパス中の話者と同じ参照表現を使うことが絶対的に正しいという仮定に基づいている。しかしながら、実際は同じ状況でも代名詞を使わない方が聞き手にとつ

てはより理解しやすい可能性もある。今後の課題としては、実際にモデルが選択した表現を人間がどれくらい正しく効率的に理解できるかを被験者実験などを通じて外生的に評価する必要があることを明らかにした。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計1件)

- (1) Philipp Spanger, Masaaki Yasuhara, Ryu Iida, Takenobu Tokunaga, Asuka Terai, Naoko Kuriyama. REX-J: Japanese referring expression corpus of situated dialogs. Language Resources and Evaluation. 2010. Dec. Online First, DOI: 10.1007/s10579-010-9134-8, 査読有.

[学会発表](計20件)

- (1) 船越孝太郎, 中野幹生, 徳永健伸, 飯田龍. 参照表現ベイジアンネットワークによる参照解決の評価. 情報処理学会 自然言語処理研究会. Vol.NL-204. No.13. pp.1-9. 2011/11/29, 石垣市.
- (2) Ryu Iida, Masaaki Yasuhara, Takenobu Tokunaga. Multi-modal Reference Resolution in Situated Dialogue by Integrating Linguistic and Extra-Linguistic Clues. Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP 2011). pp.84-92, 2011/11/9, Chiang Mai.
- (3) Naoko Kuriyama, Asuka Terai, Masaaki Yasuhara, Takenobu Tokunaga, Kimihiko Yamagishi, Takashi Kusumi. Gaze matching of referring expressions in collaborative problem solving. Proceedings of International Workshop on Dual Eye Tracking in CSCW (DUET 2011). 2011/9/25, Aarhus.
- (4) Asuka Terai, Naoko Kuriyama, Masaaki Yasuhara, Takenobu Tokunaga, Kimihiko Yamagishi, Takashi Kusumi. Using metaphors in collaborative problem solving: An eye-movement analysis. Proceedings of International Workshop on Dual Eye Tracking in CSCW (DUET 2011). 2011/9/25, Aarhus.
- (5) 栗山直子, 寺井あすか, 安原正晃, 徳永健伸, 山岸侯彦, 楠見孝. 共同問題解決における参照表現と視線一致率の関係 参照表現の属性に基づく分析. 日本認知科学会第28回大会, 2011/9/23, 東京.
- (6) 寺井あすか, 栗山直子, 安原正晃, 徳永健伸, 山岸侯彦, 楠見孝. 共同問題解決における比喩の利用 - 視線情報に基づく検討. 日本心理学会第75回大会, 2011/9/17, 東京.
- (7) 栗山直子, 寺井あすか, 徳永健伸, 山岸侯彦, 楠見孝. コンピュータを介した共同問題解決における発話同定~2者の視線一致とマウスポインタの役割の関係 ~. 電子情報通信学会 ヒューマンコミュニケーション基礎研究, 2011/8/26, 京都.
- (8) 船越孝太郎, 中野幹生, 徳永健伸, 飯田龍. ベイジアンネットワークを用いた参照表現モデルの提案. 情報処理学会 自然言語処理研究会. Vol.NL-202. No.4. pp.1-11. 2011/7/15, 山形.
- (9) 安原正晃, 飯田龍, 徳永健伸. 視線情報を利用した協調作業対話における参照解析. 言語処理学会第17回年次大会発表論文集. pp.932-935, 2011/3/10, 豊橋.
- (10) 安原正晃, 石川真也, 飯田龍, 徳永健伸. 視線情報を含むマルチモーダル協調作業対話コーパスの構築と利用. 情報処理学会自然言語処理研究会. Vol.NL-199. No.20. 2010/11/19, 広島.
- (11) 栗山直子, 寺井あすか, 安原正晃, 徳永健伸, 山岸侯彦, 楠見孝. 共同問題解決時の二者の視線一致が共有知識の形成に及ぼす効果. 日本認知科学会第27回大会, 2010/9/18, 神戸.
- (12) Tokunaga Takenobu, Yasuhara Masaaki, Terai Asuka, David Morris, Anja Belz. Construction of bilingual multimodal corpora of referring expressions in collaborative problem solving. Proceedings of the Eighth Workshop on Asian Language Resources. pp.38-46, 2010/8/23, Beijing.
- (13) Kuriyama Naoko, Terai Asuka, Yasuhara Masaaki, Tokunaga Takenobu, Yamagishi Kimihiko and Kusumi Takashi. Effect of gaze matching on collaborative problem solving. Proceedings of 7th International Conference on Cognitive Science, 2010/8/18, Beijing.
- (14) Ryu Iida, Shumpei Kobayashi, Takenobu Tokunaga. Incorporating Extra-linguistic Information into Reference Resolution in Collaborative Task Dialogue. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. pp.1259-1267, 2010/7/14, Uppsala.
- (15) Philipp Spanger, Iida Ryu, Tokunaga Takenobu, Terai Asuka, Kuriyama Naoko. Towards an Extrinsic Evaluation of Referring Expressions in Situated Dialogs. Proceedings of the Sixth International Natural Language

Generation Conference (INLG 2010).
pp.135- 144, 2010/7/9, Dublin.

- (16) 小林俊平, 飯田龍, 徳永健伸, 船越孝太郎, 中野幹生: 非言語情報を用いた協調作業対話における照応解析, 第16回言語処理学会年次大会, 2010/3/11, 東京.
- (17) 栗山直子, 寺井あすか, 安原正晃, 徳永健伸, 山岸侯彦, 楠見孝: 共同問題解決における視線の一致が解決過程に及ぼす影響, 日本認知科学会第26回大会, 2009/9/12, 藤沢.
- (18) 寺井あすか, 栗山直子, 安原正晃, 徳永健伸, 山岸侯彦, 楠見孝: 視線情報に基づく共同問題解決における比喩の効果の実験的検討, 日本心理学会第73回, 2009/8/27, 京都.
- (19) P., Spanger, Yasuhara, M., Iida, R., Tokunaga, T.: Using extra linguistic information for generating demonstrative pronouns in a situated collaboration task, Workshop on Production of Referring Expressions: Bridging the gap between computational and empirical approaches to reference, 2009/7/29, Amsterdam.
- (20) 徳永健伸, P., Spanger, 安原正晃, 飯田龍: 日本語参照表現コーパスの構築と分析, 情報処理学会 自然言語処理研究会, 2009/5/22, 東京.

6. 研究組織

(1) 研究代表者

徳永 健伸 (TOKUNAGA TAKENOBU)
東京工業大学・大学院情報理工学研究科・
教授
研究者番号: 20197875

(2) 研究分担者

栗山 直子 (KURIYAMA NAOKO)
東京工業大学・大学院社会理工学研究科・
助教
研究者番号: 90361782

寺井 あすか (TERAI ASUKA)
東京工業大学・グローバルエッジ研究院・
テニユア・トラック助教
研究者番号: 70422540