

参照表現利用に対する比喩の効果の検討

古谷 望 寺井 あすか

公立はこだて未来大学 システム情報科学部

{b1014239, aterai}@fun.ac.jp

1 はじめに

参照表現とは対象を他の対象と区別し指し示す言語表現である。一方、比喩とは対象やその特徴を他の物事を利用して示す言語表現である。どちらも対象を表現するために用いられる表現であるが、対話中においてそれぞれが互いに与える影響に関する分析はあまり行われていない。そこで本研究では、共同問題解決時における参照表現、特に指示代名詞（例：これ、それ）の利用に対する比喩の効果を明らかにすることを目的とし、対話コーパスである REX-J コーパス [1] を用いて比喩と参照表現利用の関係性を検証した。また、Spanger ら [2] の指示代名詞とそれ以外の参照表現利用を判別する SVM モデルに比喩利用に関わる情報を追加することで、比喩が指示代名詞利用に与える影響を検討した。

2 手法

2.1 コーパス

本研究では REX-J コーパス [1] を用い、参照表現における指示代名詞の利用に対する比喩の効果を検討した。REX-J コーパスは、7つのピースを組み合わせて目標の形（ゴール図形）を作るパズル課題を、操作者と解決者の二人一組で解決する際の対話を対象としたコーパスである。操作者は PC 上でパズルの操作を行えるがゴール図形が提示されず、解決者は PC 上にゴール図形が提示されるがピースを操作することができない（図 1）。二人の間にはついたりあがり、お互いの姿は見えないようになっており、PC 上に表示されている情報と対話以外で情報を共有することができない。REX-J コーパスには、パズル課題遂行中の各被験者の発話情報・参照表現に対するタグ・ピースの操作・マウスカーソルの位置・視線情報が含まれている。本稿ではこの REX-J コーパスにおける比喩表現を発話

情報から抽出し、比喩と参照表現の関係性を分析した。

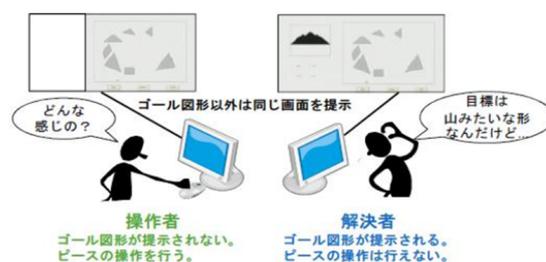


図 1: REX-J コーパスで対象としている共同問題解決 [3]

2.2 比喩の抽出

REX-J コーパスにおける発話情報に対し三者により比喩のタグ付けを行い、二者一致した表現に対し部分比喩または全体比喩のタグを付与した。部分比喩とは対象のある一部分を指し示す比喩で、全体比喩とは対象全体の形状などに関する比喩である。部分比喩と全体比喩で意見が別れたものは選択した人数の多い方の意見を採用した。また三者全員の意見が別れた表現は比喩ではないと判断した。

実験参加者: 大学生・大学院生・社会人各 1 名

材料: REX-J コーパス (8 組 31 対話)

手続き: REX-J コーパスの発話情報を確認し、比喩だと思われる表現に部分比喩または全体比喩のタグ付けを行う。具体的にはその表現が含まれる「発話開始の対話開始からの時刻」「発話終了の対話開始からの時刻」「比喩だと思われる表現」「発話者」「発話中の比喩だと思われる表現」を記録する。さらに、その比喩表現が部分比喩、または、全体比喩のどちらと思われるかを記録する。この作業を 31 対話分、三者がそれぞれ行い、比喩分類表 (表 1) を作成した。この 3 人分

表 1: 比喩分類表

発話の開始時間	発話の終了時間	発話	発話者	比喩表現	比喩の種類
00:05.030	00:06.130	なんかキリンみたいななの。	解決者	キリン	全体
00:06.340	00:06.700	キリン？	操作者	キリン	全体
00:07.040	00:09.100	で、右、右側に首が来るんだけど。	解決者	首	部分
00:09.320	00:09.980	右に首ね。	操作者	首	部分
00:14.820	00:15.160	あたま	操作者	あたま	部分

の比喩分類表から三者とも共通して比喩だと認識している表現に対し部分比喩または全体比喩のタグを付与した。全体比喩か部分比喩かについては三者のうち二者以上が判断している種類とし、三者全員の意見が別れた表現は比喩ではないと判断した。

3 分析結果

3.1 対話における比喩の有無と参照表現利用の関連

本研究で利用したコーパス 31 対話のうち、パズル課題遂行中に比喩が使用された対話は 23 対話であった。全対話から参照表現が使用された回数は 2128 回であった。そのうち、指示代名詞は 978 回、それ以外の参照表現は 1150 回であった。比喩が使用された対話とそうでない対話の、参照表現別の使用された回数を表 2 に示す。カイ二乗検定を用いて分析を行ったところ、比喩が使用された対話ではそうでない対話と比較し、指示代名詞の使用が 1%水準で有意に多く、それ以外が有意に少なく比喩の利用が参照表現利用における、指示代名詞の利用を促進している可能性が示唆された ($\chi^2(1) = 52.36$)。

表 2: 対話における比喩の有無と参照表現の使用回数

	指示代名詞	それ以外	合計
比喩使用	923	971	1894
比喩未使用	55	179	234
合計	978	1150	2128

3.2 比喩利用の前後における参照表現利用

比喩の利用による、指示代名詞の利用の促進についてより詳細に検討するため、比喩利用の前後における参照表現別の出現回数を調べた。使用したコーパスから抽出した合計 462 の比喩 (部分比喩: 347, 全体比喩: 115) に関し、直前での参照表現別の使用回数を表 3、比

喩の直後での参照表現別の使用回数を表 4 に示す。カイ二乗検定を用いて分析した結果、比喩の直前では参照表現の違いによる有意差は見られなかったが、部分比喩と比較して全体比喩の直前では「なし」の割合は 1%水準で有意に高く、参照表現が有意に使用されていなかった ($\chi^2(2) = 92.17$)。一方で、比喩の直後では、全体比喩と比較し部分比喩の直後において指示代名詞が 1%水準で有意に高く、それ以外が有意に低いという結果であった。このことから、全体比喩利用の直前での参照表現の利用が少なく、全体比喩と比較し、部分比喩の直後に指示代名詞利用が促進される可能性が示唆された ($\chi^2(2) = 10.99$)。

表 3: 比喩利用の直前の参照表現別の使用回数

	指示代名詞	それ以外	なし	合計
部分比喩	182	146	19	347
全体比喩	41	26	48	115
合計	203	172	67	462

表 4: 比喩利用の直後の参照表現別の使用回数

	指示代名詞	それ以外	なし	合計
部分比喩	162	180	5	347
全体比喩	34	80	1	115
合計	196	260	6	462

3.3 比喩利用と参照表現利用の時間間隔

比喩利用の前後における参照表現までの時間について分析を行った (図 2,3)。その結果、部分比喩と直前の指示代名詞までの平均間隔 24.6 秒、それ以外の参照表現までの平均間隔 23.6 秒、全体比喩と直前の指示代名詞までの平均間隔 27.8 秒、それ以外の参照表現までの平均間隔 34.1 秒であり、全体比喩と比較し部分比喩は直前の参照表現までの時間間隔が 5%水準で有意に短いことが示された ($F(1, 372) = 5.313$)。しかし、指示代名詞とそれ以外の参照表現では、有意差は

見られなかった。部分比喩の直後の指示代名詞までの平均間隔 27.9 秒、それ以外の参照表現までの平均間隔 23.6 秒、全体比喩の直前の指示代名詞までの平均間隔 37.3 秒、それ以外の参照表現までの平均間隔 36.3 秒であり、部分比喩の方が全体比喩よりも参照表現までの時間が 0.1%水準で有意に短いことが明らかになった ($F(1, 446) = 17.53$)。しかし、指示代名詞とそれ以外の参照表現では、有意差は見られなかった。この結果から、比喩の種類によって参照表現利用との近接性に違いがあることが示唆されたが、参照表現の種類（指示代名詞またはそれ以外）に与える影響は小さいことが示唆された。

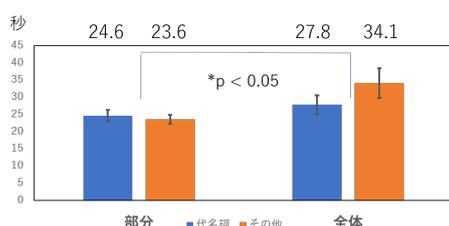


図 2: 比喩と直前の参照表現までの平均間隔

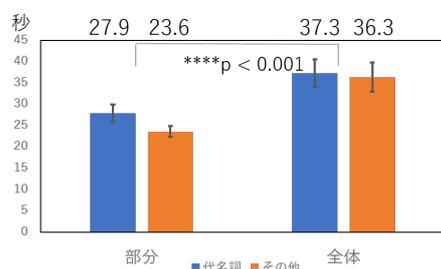


図 3: 比喩と直後の参照表現までの平均間隔

4 SVM モデルを用いた検討

Spanger ら [2] は、直前の課題遂行履歴から指示代名詞とそれ以外の参照表現を判別する SVM モデルを構築した。この SVM モデルは、ある対象が参照された状況に基づき、使用された参照表現が指示代名詞かそれ以外の参照表現かを判別する SVM モデルである。Spanger らは、この SVM モデルへの入力データとして、対象が参照された状況を表す特徴量を 12 種定義し、これらの特徴量を談話履歴に関する特徴量 (D1~D5)、操作履歴に関する特徴量 (A1~A5)、および現在の操作に関する特徴量 (O1, O2) の 3 つのカテゴリに

分類した。各特徴量はその値によって定義された。たとえば、時間的距離を表す特徴量 (D1, D4, A1, A4) は時間適距離を t とすると “ $t < 10 (< 10)$ ”, “ $10 \leq t < 20 (< 20)$ ”, “ $20 \leq t < \infty (< \infty)$ ” 秒の 3 水準に分類された (表 5)。

本稿では Spanger らの SVM モデルに比喩に関する特徴量を加えることで、比喩が参照表現に与える影響をシミュレーションにより検証した。加えた特徴量は、「最後に使用された比喩の種類 (部分/全体)」と「最後に比喩が使用されてからの時間 ($< 10, < 20, < \infty$ または $< 28, < 56, < \infty$)」である。「最後に比喩が使用されてからの時間」は、Spanger ら [2] の定義を参考にした ($< 10, < 20$) と、比喩と直後の指示代名詞とそれ以外の参照表現の平均時間間隔 (指示代名詞: 平均 30.2 秒、それ以外の参照表現: 平均 26.7 秒) から、2 つの平均時間間隔に基づく (< 28) とその 2 倍の (< 56) を用いた。

本稿では、この合計 21 種の特徴量を用いて、指示代名詞とそれ以外の参照表現を判別する SVM モデルを構築し、10 分割の交差検証を用いてシミュレーションを行うことで、特徴量の有無とモデルの精度の関連を検討した (表 6)。一番左の行に使用された特徴量の組み合わせが示されている。「All(10, 20)」は表 5 のすべての特徴量と、最後に使用された比喩の種類、最後に使用された比喩との時間 ($< 10, < 20$) を使用することを意味し、「All(28, 56)」は表 5 のすべての特徴量と、最後に使用された比喩の種類、最後に使用された比喩との時間 ($< 28, < 56$) を使用することを意味する。「w/oDx」は談話履歴 (Dx) を除くすべての特徴量を使用すること意味し、「w/oAx」, 「w/oOx」, 「w/oMx」はそれぞれの操作履歴、現在操作、比喩履歴に関わる特徴量を除くことを意味する。この結果、「All(28, 56)」の場合に、パフォーマンスが最も良いことが分かる。また、現在操作に関する特徴量を除いた場合 (w/oOx) と比較し、比喩履歴を除いた場合 (w/oMx) に F-measure がより低下することから、現在操作より比喩利用履歴が指示代名詞利用に影響を与えている可能性が示された。さらに、「All(10, 20)」と比較し「All(28, 56)」は Precision・Recall ともに上昇することから、最後に比喩を使用してからの時間が指示代名詞利用に影響を与えている可能性が示された。

5 おわりに

本研究では、参照表現における指示代名詞の利用に対する比喩の影響を検討した。その結果、比喩が使用

表 5: 特徴量表

談話履歴に関する特徴量		
D1	< 10, < 20, < ∞	対象が最後に参照されてからの時間
D2	yes, no	対象が最後に参照されたとき, 指示代名詞が使用されたか否か
D3	integer	対象が最後に参照されてから, 他のピースが参照された回数
D4	< 10, < 20, < ∞	他のピースが最後に参照されてからの時間
D5	yes, no	最後に参照されたのが対象か否か
操作履歴に関する特徴量		
A1	< 10, < 20, < ∞	対象が最後に操作されてからの時間
A2	flip, move, rotate	対象が最後にされた操作の種類
A3	integer	対象が最後に操作されてから, 他のピースが操作された回数
A4	< 10, < 20, < ∞	他のピースが操作されてからの時間
A5	yes, no	最後に操作されたのが対象か否か
現在の操作に関する特徴量		
O1	yes, no	参照が始まったとき, 対象が操作中だったか
O2	yes, no	参照が始まったとき, 対象がマウスカーソルの下にあったか
比喩に関する特徴量		
M1	< 10, < 20, < ∞ < 28, < 56, < ∞	最後に比喩を使用してからの時間
M2	par, all	最後に使用された比喩の種類

表 6: シミュレーション結果

Features	Precision	Recall	F-measure
All(10, 20)	0.7407	0.5491	0.6307
All(28, 56)	0.7524	0.5532	0.6376
w/o Dx	0.7390	0.4663	0.5718
w/o Ax	0.6329	0.5235	0.5730
w/o Ox	0.7482	0.5409	0.6279
w/o Mx	0.7433	0.5419	0.6268

された対話ではそうでない対話よりも, 指示代名詞の使用が有意に多く, それ以外の参照表現が有意に少ないことが明らかになった. このことから, 比喩の利用が参照表現利用における, 指示代名詞の利用を促進している可能性が示唆された. また, 直前に使用する参照表現に指示代名詞を選択するか否かの判断に比喩の違いはあまり影響がないが, 直前に参照表現を使用するか否かの判断には影響があることが示唆された. 全体比喩利用の直前での参照表現の利用が少なく, 全体比喩と比較し, 部分比喩の直後に指示代名詞利用が促進されるとともに, 部分比喩の方が全体比喩よりも直前・直後の参照表現までの時間が有意に短く, 比喩の種類によって参照表現利用との近接性に違いがあることが示唆された. さらに SVM を用いたモデルシミュレーションの結果, 比喩利用履歴が指示代名詞またはそれ以外の判別と一定程度の関わりがあり, 現在操作より比喩利用履歴が指示代名詞利用に影響を与えてい

る可能性が示された.

参考文献

- [1] Spanger, P., M. Yasuhara, R. Iida, and T. Tokunaga: “REX-J: Japanese referring expression corpus of situated dialogs”, Language Resources and Evaluation, v.46 n.3, p.461-491, September 2012
- [2] Spanger, P., M. Yasuhara, R. Iida, and T. Tokunaga. (2009): “Using extra linguistic information for generating demonstrative pronouns in a situated collaboration task”
- [3] 栗山・寺井・安原・徳永・山岸・楠見 (2011). “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-”