

単語分散表現を用いたロボットのジェスチャー生成

Generating robot gesture from distributed word representation

佐々木康佑¹ 西川純平¹ 森田純哉¹

Alexis Meneses² 酒井和紀² 吉川雄一郎²

Kosuke Sasaki¹ Jumpei Nishikawa¹ Junya Morita¹

Alexis Meneses² Kazuki Sakai² Yuichiro Yoshikawa²

¹ 静岡大学 ² 大阪大学

¹ Shizuoka University ² Osaka University

Abstract: This study proposes a method of generating robot gestures from distributed word representations. In the method, the size image for words is computed based on the axis whose poles correspond to “small” and “large” words images. In addition, the size image of the words is physically implemented as robot gestures. The proposed methods was evaluated by two online surveys. Summarizing the results, the authors claim the possibility of exchanging qualitative and quantitative aspects of word representations.

1 はじめに

人間と機械のインタラクションは、記号もしくは数量的な表現に媒介される。記号的なインタフェースでは、言語やアイコンなどの離散的な表現が用いられ、数量的なインタフェースでは音声や動作などの物理量が用いられる。記号と数量の2つの媒体は、人間内部においても異なるシステムにより相補的に処理される [1].

このような議論は認知科学や人工知能分野において以前より行われている。ホーキンスの提唱した座標系理論によると概念の裏側には概念に関する知識を内包する連続的な空間があり、その知識は言語などを介して復元され、身体動作などに変換されるとされている [2]. また、そのような時空間に埋め込まれた知識構造をベースとして思考が形成され、思考によって言語が形成される [3].

そのような記号と数量的な表現を変換する仕組みを考える上で、単語の意味に関するモデルが重要となる。人間の言語利用において、単語の意味は離散的・独立に定義されるのではなく、単語同士が相互に関連付けられた連続的な空間の中で定められると考えられる。自然言語処理に関する研究では、こういった単語間の意味的な関係を捉えるために、人間の言語運用から得られたコーパスに対して統計的な解析 (Bag of Words, 共起頻度の計算, あるいは単語ベクトルからの主成分分析) が適用されてきた。さらに近年では、ニューラルネットの中間層に畳み込まれたベクトル表現 (単語分散表現) を用いることで、単語の数量的意味を把握することが主流になっている。

上記の背景から、本研究では離散的な記号と数量的イメージの変換機構を利用した身体的イメージの生成手法を提案する。この手法では、ニューラルネット上に構築された単語分散表現から、単語の背後にある空間的な知識構造を抽出する。冒頭の理論的背景 [1, 2, 3] に従えば、単語が位置づけられる空間には、「大きさ」や「速さ」など様々な数量的なイメージが存在すると想定できる。本研究では、それらのうち「大きさイメージ」に着目し、大きさに関する単語の数量表現を得る。本研究では人間がそれぞれの単語に持つ、一般的な大きさのイメージを単語の「大きさイメージ」と定義する。そして、その表現を身体動作に変換することで、ロボットのアイコンニックジェスチャー (身体的イメージ) を生成する。

本研究は以下の仮説を検証することを目的とする。

1. 単語分散表現に内包される空間的な知識構造が「大きさイメージ」を含有している
2. 単語分散表現から生成した「大きさイメージ」によって生成した身体的イメージは人間の認識を復元できる

以下、本研究の目的につながる背景として、2章では本研究と関連する単語分散表現やロボットのジェスチャー生成に関する研究をレビューする。3章では本研究で提案する概念の身体的イメージの生成手法を述べる。4章ではアンケート調査により得られた「大きさ」に関する人間の認識と整合する「大きさイメージ」を構築し、1つ目の仮説を検証する。5章では「大きさ

イメージ」により生成したアイコンックジェスチャーに関する主観評価を実施し、2つめの仮説を検証する。6章では本研究のまとめと今後の展望を示す。

2 関連研究

単語、あるいは概念の意味は、複数のアプローチによりモデル化される。一つは、社会で流通する概念の意味を人手によって書き出すものである。これまで、WordNetやConceptNet、Cycなど大規模なデータベースが整備されている。これらは社会における概念の規範的な知識構造を定義するものといえる。

他方で、近年では人間が日常的に用いる言語の用法から概念の意味を統計的に捉えるアプローチが盛んである。単語分散表現 [4] は、単語をベクトル空間に埋め込まれた1つの点とみなす。この枠組みにおいて、単語の意味は、ベクトル空間上での単語間の関係（距離あるいは類似）と捉えられる。その背景にある考え方は、「単語の意味は周囲の単語によって決定され、意味の類似した単語は類似した文脈で使用される」という分布仮説 [5] である。

単語分散表現を用いることで、単語の数量的なイメージを抽出する試みがなされている。たとえば、内海 [6] は単語分散表現に埋め込まれた内部知識を明らかにすることに取り組んだ。単語分散表現を利用することで単語を属性分類し、その結果と人間に対する調査から得られた単語の属性分類を比較した。結果、人間の持つ知識が単語分散表現のベクトル空間に内包されていることが示唆された。中でも物理的・身体的な意味を持つ単語に対して、抽象的な意味を持つ単語において、その属性が単語分散表現に埋め込まれていることが示された。

Grandら [7] は、単語分散表現を用いて文脈依存の関係を抽出する手法を提案している。文脈依存の関係というのは、「犬」という単語でも文脈によって「大きさ」「知能」「危険性」など複数の異なる意味的特徴を持つという事を表す。この研究においては、単語ベクトルを「大きさ」「知能」「危険」などの特徴を表す軸に射影することで様々な物体のカテゴリや特性に対する人間の判断を復元できることが示された。

このように、単語分散表現には人間の持つ単語の知識（数量的イメージ）が埋め込まれていることが示唆されている。すなわち、単語分散表現には人間が身体的に獲得した、概念の定量的意味が内在していることが分かる。

こういった概念の定量的な意味がコミュニケーションに与える影響とはどのようなものだろうか。この点に関して、下川原ら [8] は発話内容の感情に合わせたジェスチャーをロボットに行わせる手法を提案し、人

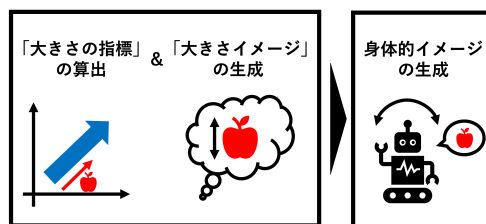


図 1: 提案手法の流れ

間にどのような影響を与えるのかを検証している。また、福原ら [9] はロボットによるジェスチャーが行われるタイミングに着目している。この研究においてはジェスチャーを発話内の適切なタイミングで行うことで人間がロボットをコミュニケーションパートナーとして認識するという結果を示している。

このようにロボットのジェスチャー生成に関する研究は数多く存在する。しかし、単語分散表現を用いて獲得した概念の定量的意味をロボットのアイコンックジェスチャー（身体的イメージ）の生成に結びつけた先行研究は見つけられていない。よって、本研究では単語分散表現のベクトル空間から「大きさ」に関する概念の定量的意味を獲得し、身体的イメージを生成する手法を提案する。さらに、生成された身体的イメージに対する人間の評価を収集することで、人間の認識との整合性を検証する。

3 提案手法

本研究の提案手法は、図1のステップを含む。以下、各ステップの枠組みを示す。

3.1 「大きさイメージ」の生成

1節にて、単語分散表現のベクトル空間には「大きさ」に関する数量的イメージが埋め込まれているという仮説を提起した。この仮説に従って「大きさイメージ」を生成するためには、単語のベクトル空間における「大きさ」の軸（指標）を定義する必要がある。以下、多次元ベクトル空間から「大きさの指標」を抽出し、任意の単語に対する「大きさイメージ」を取り出すまでの処理を示す。

1. 「大きさの指標」の構成

単語分散表現の多次元ベクトル空間から単語の大きさの指標を構成する手法は、Grandら [7] に従う。この手法では、大きさの指標を「大きい」と「小さい」の両極からなる軸上の値ととらえる。

両極を定義するためには、分散表現空間上の「大きい」と「小さい」の座標を抽出すればよい。しかし、この2つの単語には、大きさに関する意味のほかに、形容詞であることに由来する余剰な意味が備わる。そういった大きさと関係しない意味を除外するために、単語分散表現の中で、「大きい」「小さい」と同様の役割を有する類義語集合を定義する。そして、それら類義語集合の平均ベクトルを計算することで極座標を定める。

「大きい」と「小さい」の類義語集合は人力で構成されたシソーラスを参照することで定めることができる。通常、単語は多義的であり、複数の意味を有する。単語の類義語は、それらの意味ごとに集合 (synset) として定義される。これらの synset から、人間の認識と整合する極座標を定めるために、人間に対する調査から得られる大きい単語、小さい単語の距離を最大化させる synset の組み合わせを求める。

この抽出において、単語が指し示す対象のカテゴリを考慮する。Tversky [1] などによれば、概念の意味はもともと生物の動作から構成される。しかし、内海 [6] が示すように、社会的に流通する文書から構成された単語分散表現のなかで、生物的数量は強くは埋め込まれないと予測される。

2. 「大きさイメージ」の構成

「大きさの指標」から入力単語ベクトルの類似度を算出し、その値を入力単語に対する「大きさイメージ」とする。本研究の提案手法においては、この数値、いわゆる「大きさイメージ」が大きいほど、「大きい」単語、数値が小さければ小さいほど「小さい」単語と認識される。

3.2 身体的イメージの生成

各単語に対して生成された「大きさイメージ」の値に即して、身体的イメージ (アイコンニックジェスチャー) を物理化する。パラメータのスケーリングと手順は以下に示す。

1. 大きい姿勢と小さい姿勢の設定

「大きさの指標」を身体の部位によって構成される姿勢とマッピングする。そのために、人間が認識する単語のなかで最も小さい単語、最も大きい単語と対応する身体姿勢を定める。この姿勢を基準 (最も小さい単語のイメージを0, 最も大きい単語のイメージを1) として、各単語の「大きさイメージ」を0から1の範囲に位置付ける。

2. 各関節におけるパラメータの計算

姿勢を構成する各関節の関節角度に対して、上記のスケーリングを適用する。

3. 身体的イメージの生成

2により得られた値をもとにジェスチャーを生成する。なお、ジェスチャーは単語の発話と同時にされるものとする。

4 実験1: 「大きさイメージ」の生成

提案手法のうち「大きさイメージ」の生成を行うことで、1節に示した1つめの仮説を検証する。単語の大きさに関する人間の認識を抽出するアンケート調査を実施し、その結果を用いて、「大きさの指標」を生成する。

4.1 方法

4.1.1 分散表現モデル・シソーラス

本研究では分散表現モデルとして、鈴木ら [10] によって開発された日本語 Wikipedia エンティティベクトルを用いた。このモデルは、word2vec [11] によって構築された学習済みモデルである。word2vec は、BERT [12] など、より近年に開発された分散表現モデルに比べて軽量であり、リアルタイム性を求められる身体イメージの生成に有効と考えた。

さらに、本研究では、類義語の選定に日本語 WordNet [13] を用いた。日本語 WordNet には「大きい」に対して28個、「小さい」に対して14個の synset が含まれている。これらの synset から日本語 Wikipedia エンティティベクトルに含まれない単語は削除した。また、属する類義語の存在しない synset を削除した。結果、「大きい」の synset は23個、「小さい」の synset は13個となった。これら $23 \times 13 = 299$ 通りに対して、「大きさの指標」を算出した。

4.1.2 アンケート調査

人間の認識と整合する「大きさの指標」を定義するために、オンラインアンケートによって、人間が認識する大きい単語と小さい単語の集合を取得した。アンケート回答者は、クラウドソーシングサイト Lancers にて募集した100名であった。単語のカテゴリとして、生物、無生物、無形概念を設定し、表1の設問を用意した。回答者は、設問の回答を自由記述で記載し、謝金は55円を受け取った。

表 1: アンケートの設問

	設問
1	あなたが大きいと思う生物（人間・動物・植物など）を 5 個挙げて下さい。
2	あなたが小さいと思う生物（人間・動物・植物など）を 5 個挙げて下さい。
3	あなたが大きいと思う無生物（自然物・人工物など）を 5 個挙げて下さい。
4	あなたが小さいと思う無生物（自然物・人工物など）を 5 個挙げて下さい。
5	あなたが大きいと思う、形状を持たない概念（名詞）を 5 個挙げて下さい。
6	あなたが小さいと思う、形状を持たない概念（名詞）を 5 個挙げて下さい。

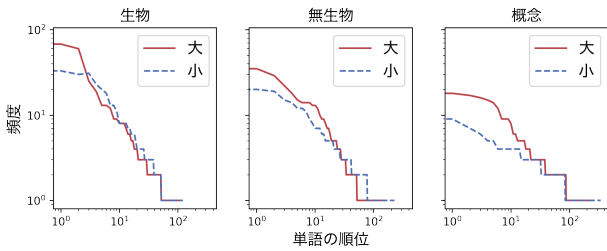


図 2: アンケートにより得られた回答の順位頻度

4.2 結果

4.2.1 アンケート回答の分布

本調査により得られた単語数は 3 つのカテゴリを合わせて 937 単語であり、そのうち日本語 Wikipedia エンティティベクトルに含まれる単語数は 828 単語であった。これらの単語のうち、大きい生物が 62 単語、無生物が 117 単語、無形概念が 188 単語、小さい生物が 83 単語、無生物が 150 単語、無形概念が 228 単語であった。図 2 は、これらの単語の順位頻度を両側対数グラフとして示している。また、表 2 には各設問に対して回答された上位 5 単語とその頻度を示す。

4.2.2 「大きさの指標」の抽出

299 の「大きさの指標」(synset の全組み合わせと対応) から、人間の認識と最も整合する指標を選定する。この選定において、大きい単語の「大きさイメージ」、小さい単語の「大きさイメージ」の平均値をカテゴリごとに計算し、その差分を求める。図 3 は、299 の指標に対して計算された差分を示している。この図の横軸は、順位ごとに並べられた synset の組み合わせに対応する。この図から、全カテゴリに共通して一部の組み合わせの「大きさイメージ」の差分が大きくなっていることが分かる。また、全体的に見ると「大きさイメージ」の差分が 0 より大きい組み合わせ（赤い点線より上）が多く、synset の組み合わせの半数以上において、人間の認識と整合する大きさの指標が計算されることが示された。

表 2: 各アンケート設問に対して得られた回答（上位 5 単語）

(a) 生物

大		小	
単語	頻度	単語	頻度
ゾウ	85	アリ	74
クジラ	68	ミジンコ	33
キリン	60	蚊	31
クマ	25	ダニ	30
カバ	19	ノミ	23

(b) 無生物

大		小	
単語	頻度	単語	頻度
東京スカイツリー	45	砂	26
富士山	35	ピーズ	20
東京タワー	29	針	26
エベレスト	22	マイクロチップ	15
ピラミッド	18	ネジ	14

(c) 無形概念

大		小	
単語	頻度	単語	頻度
宇宙	36	心	13
愛	18	嫉妬	9
夢	17	妬み	7
心	16	見栄	6
海	15	点	5

図 4 には 3 つのカテゴリの順位を合計した際の上位下位 3 つの差分を示す。上位の組み合わせにおける「大きい」の synset として“outstanding”, 「小さい」の synset として“flyspeck”, 下位の組み合わせにおける「大きい」の synset として“sizable”, 「小さい」の synset として“immature”が多く現れていた。これらより「大きい」「小さい」どちらも共通して、抽象的な意味を持つ synset が上位となり、生物に関係する意味を有する synset が下位となる傾向があったといえる。

そして、3 つのカテゴリの総合順位の最も小さい（平均して最も差分の大きい）synset の組み合わせは“larger-than-life”と“flyspeck”であった。表 3 に示すように、“larger-than-life”は「非常に印象的である、普通に勝る」という意味を持ち、“flyspeck”は「非常に小さい」という意味を持つ。そして、それぞれに「壮大、大規模」、「ちっぽけ、微小、ちっこい、ごく小さい」という類義語が属している。これらのうち“flyspeck”に属する、「ちっこい、ごく小さい」という単語は日本語 Wikipedia

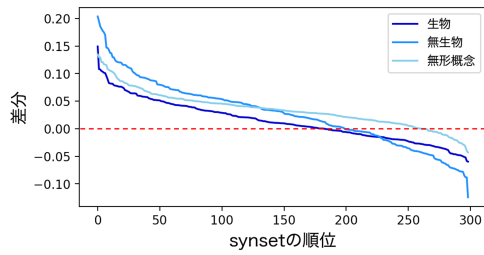


図 3: 「大きさイメージ」の差分の分布

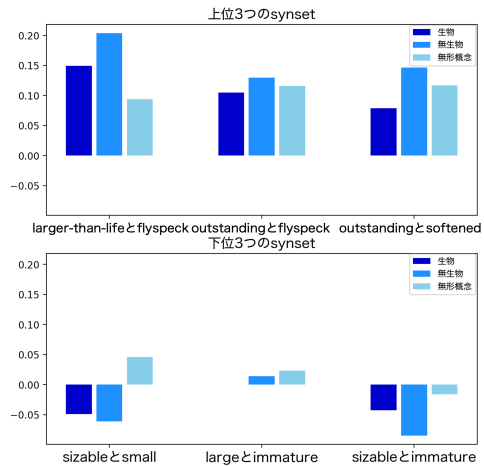


図 4: 差分の合計順位の上位 3 つ, 下位 3 つの synset の組み合わせ

エンティティベクトルにおいて未学習であったため、本研究における「大きさの指標」の算出においては除外した。

4.2.3 「大きさイメージ」の検証

4.2.2 の検討によって抽出された“larger-than-life”と“flyspeck”を用いた「大きさの指標」により計算される「大きさイメージ」を検討する。図 5 は、アンケートにより得られた各カテゴリに含まれる単語の「大きさイメージ」の平均値である。図から、全てのカテゴリにおいて「大きい」単語の「大きさイメージ」が「小さい」単語の「大きさイメージ」を上回っていることが見て取れる。この印象を確かめるため、「大きさイメージ」を従属変数と

表 3: 「larger-than-life」と「flyspeck」の単語一覧

単語	「大きい」	「小さい」
synset	larger-than-life	flyspeck
meaning	非常に印象的である	非常に小さい
類義語 1	壮大	ちっぽけ
類義語 2	大規模	微小
類義語 3	ちっこい	
類義語 4	ごく小さい	

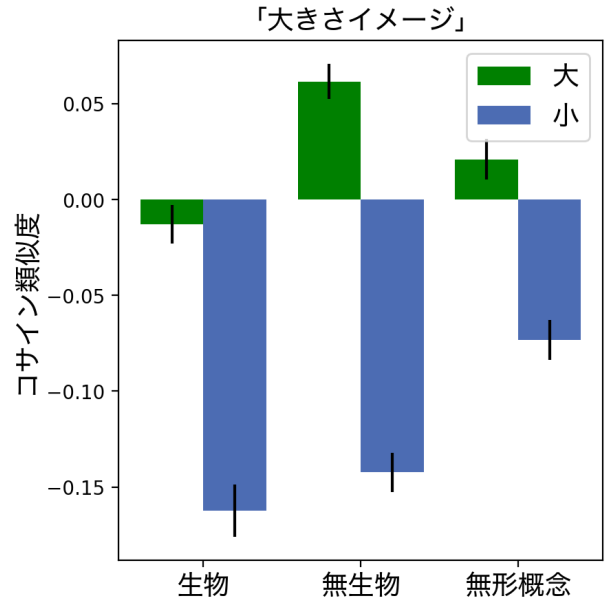


図 5: アンケートにより得られた単語群と「大きさの指標」とのコサイン類似度（大きさイメージ）の平均値（以降、エラーバーは標準誤差）

して 2 要因 [カテゴリ (生物, 無生物, 無形概念) × 質問におけるサイズ (大きい, 小さい)] 分散分析 (ANOVA) を行った。結果、要因間の交互作用 ($F(2, 824) = 8.84, p < .001$) が有意となり、生物 ($F(1, 824) = 65.60, p < .001$), 無生物 ($F(1, 824) = 122.14, p < .001$), 無形概念 ($F(1, 824) = 26.06, p < .001$) におけるサイズの単純主効果が有意となった。

以上より、全てのカテゴリにおいて「大きい」単語の「大きさイメージ」は「小さい」単語の「大きさイメージ」より大きいことが確かめられた。つまり、“larger-than-life”と“flyspeck”を用いて算出した「大きさの指標」によって、人間の認識と整合する単語の大きさを判別できることが示された。

4.3 考察

上記より、人間の認識する大きい単語と小さい単語を区別する軸が、単語分散表現のベクトル空間上に定義できた。つまり、本研究の 1 つめの仮説「単語分散表現のベクトル空間には「大きさ」に関する数量的イメージが埋め込まれている」が支持された。

ただし、この結果は、アンケートから得られた大きい単語と小さい単語の差分を最大化する軸を定めただけでなく、驚くべきものではない。興味深い結果は、図 3 の上位において抽象的な意味を持つ synset, 下位において生物に関する意味を持つ synset が観察されたことである。この結果は、単語分散表現のベクトル空間

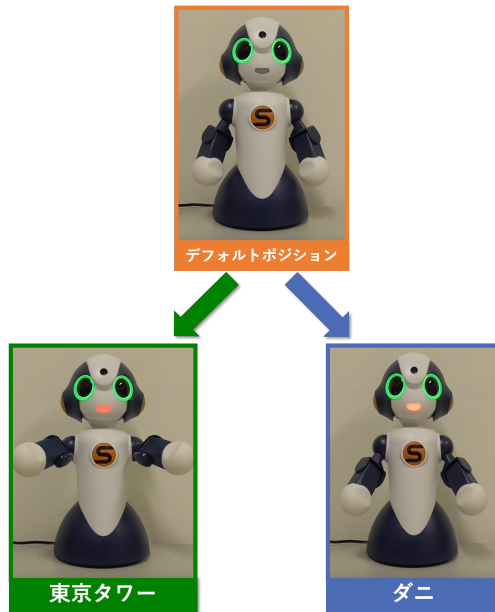


図 6: Sota のジェスチャー例 (左:「東京タワー」、右:「ダニ」)

には抽象的な意味が多く内包されているという指摘 [6] を支持するものである。

5 実験 2: 「身体的イメージ」の生成

前節で生成した「大きさイメージ」に即してロボットの身体イメージ (アイコンニックジェスチャー) を生成し、そのイメージに対する評価を行う。この検討により、1 節に示した 2 つめの仮説を検証する。

5.1 方法

5.1.1 機材・材料

「身体的イメージ」の生成 身体イメージの物理化には、Vston 社による小型コミュニケーションロボット Sota を用いる。Sota の身体動作は 9 つの関節 (3 つの首の関節, 左右 2 つの肩の関節, 左右 2 つの腕の関節, 1 つの胴体の関節) によって制御される。これらの関節の角度や動かすスピードを制御することで様々な動きを生成できる。さらに、Sota は発話機能を持ち、自由な単語を発話させることができる。

本研究では、腕と肩の関節のパラメータを「大きさイメージ」により定めることでジェスチャー生成を行った。Sota のデフォルトの姿勢として図 6 の上部の画像にあるような肩を下ろして腕を少し曲げている状態を設定した。この状態から腕と肩の関節のパラメータを

表 4: 「大きさイメージ」、各関節のパラメータの最大値・最小値

	最大値	最小値
大きさイメージ	0.310	-0.490
肩	27	-70
腕	-25	20

変化させることで単語の大きさと対応したジェスチャーを生成する。発話する単語の「大きさイメージ」が大きい時には肩を上げて腕を広げ、小さい時には腕・肩どちらもあまり動かさない。

表 4 に、調査により得られた全単語で算出した「大きさイメージ」と、肩と腕の関節のパラメータの最大値・最小値を示す。この値を用いて、「大きさイメージ」の値が 0 から 1 となるようにスケーリングした。

生成された身体的イメージの例を図 6 で示す。左下の画像が「東京タワー」(大きい無生物) と発話した際のジェスチャー、右下の画像が「ダニ」(小さい生物) と発話した際のジェスチャーを示している。

アンケート 表 2 の 30 単語を対象に身体的イメージを生成し、動画を作成した。この動画に対して、ロボットの動作と発話している言葉の対応が自然であるかを 5 段階で評価させる (1: 全く自然でない—5: 非常に自然である)。「自然さ」とはロボットとの対話を想定したものを指し、具体的には「ロボットが人間のように言葉の意味を理解していると感じるかどうか」と教示した。

また、「大きさイメージ」によって生成した身体的イメージの評価との比較対象として逆転条件を設定した。逆転条件では、4.2.1 のスケーリングとは反対に、スケーリングされた値が 1 であるときに各関節の最小値を取り、0 であるときに各関節の最大値を取る。これにより、「大きさイメージ」の値が大きいときには腕と肩があまり動かされず、小さいときには腕を上げて大きく開かれるようになる。

また、順序効果を防ぐために、大きさイメージ条件と逆転条件の 60 の動画が無作為に並び替えた 3 つのフォームを用意した。調査対象者はアクセスした順にこれらのフォームに均等に振り分けられた。また、60 の動画の評価の合間に、指定の数字を回答させるダミー質問を 2 問設定した。

5.1.2 アンケート調査

Lancers にて依頼を受けた参加者 200 名 (男性 133 名, 女性 62 名) は、依頼画面に示される教示に目を通したうえで調査に参加した。教示では、評価手順、自然さの定義、ダミー質問の回答義務が説明された。教

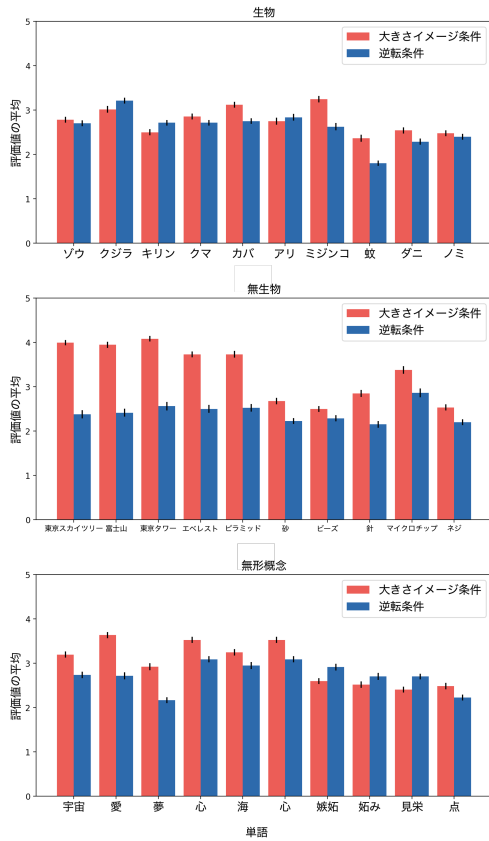


図 7: 単語ごとの評価値の平均

示に同意したのちに、参加者は Google フォーム上で、62 問の質問に回答した。アンケートの回答後には ID が表示された。表示された ID をランサーズの依頼画面に入力することで、依頼が完了し、参加者は謝礼 110 円を受け取った。

5.2 結果

複数回の回答が記録された 2 名、ダミー質問の回答を誤った 2 名の回答を削除した。分析対象となった 196 の回答に対して、単語別の評価値の平均を計算した (図

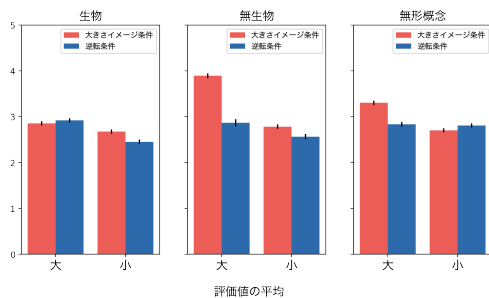


図 8: カテゴリごとの評価値の平均

表 5: 大ききイメージ条件と逆転条件の単純主効果

カテゴリ	サイズ	平均値差	F 統計量	p 値	効果量 d
生物	大きき	-0.067	4.85	0.029	0.103
	小さい	0.222	28.19	<.001	0.328
無生物	大きき	1.031	157.61	<.001	1.107
	小さい	0.219	15.80	<.001	0.299
無形概念	大きき	0.466	70.16	<.001	0.652
	小さい	-0.104	7.13	0.008	0.165

効果量 d の基準: $d = 0.200$ (効果量小), $d = 0.500$ (効果量中), $d = 0.800$ (効果量大)

7). さらに、これらの単語をカテゴリごとに総合し、評価値の平均を計算した (図 8). この値を従属変数とし、3 要因 [条件 (大ききイメージ条件, 逆転条件) \times サイズ (大きき, 小さい) \times カテゴリ (生物, 無生物, 無形概念)] 分散分析 (ANOVA) を行った. その結果、条件 ($F(1, 195) = 93.31, p < .001$) の主効果が認められ、全体としては大ききイメージ条件が逆転条件よりも自然な評価を得たことが確かめられた. しかし、2 次の交互作用 ($F(2, 390) = 86.97, p < .001$) も有意となり、大ききイメージと逆転条件の差の現れ方が、カテゴリやサイズによって異なることが示された.

各カテゴリにおける条件の差を検討するために、カテゴリ (生物, 無生物, 無形概念) ごとに 2 要因 [条件 (大ききイメージ条件, 逆転条件) \times サイズ (大きき, 小さい)] 分散分析 (ANOVA) を行った. その結果、いずれの分散分析においても要因間の交互作用が有意となった (生物: $F(1, 195) = 32.01, p < .001$), 無生物: $F(1, 195) = 89.14, p < .001$, 無形概念: $F(1, 195) = 103.74, p < .001$).

表 5 は「大きき」と「小さい」の各サイズ要因ごとに、条件の単純主効果 (大ききイメージ条件-逆転条件) を示している. 表より、生物における小さい単語、無生物における大きき単語と小さい単語、無形概念における大きき単語において、大ききイメージ条件の自然さが逆転条件の自然さを上回ったことがわかる. これらのうち無生物における大きき単語、無形概念における大きき単語は、効果量が中程度以上であり、自然なジェスチャーの生成がなされたといえる. しかし、小さい効果量ではあるものの、生物における大きき単語、無形概念における小さい単語において、逆転条件の自然さが大ききイメージ条件の自然さを上回った.

5.3 考察

上記の結果、全体的な自然さの評価は、大ききイメージ条件が逆転条件を上回った. また個別の検定においても、多くのカテゴリのサイズにおいて、自然さの評価は大ききイメージ条件が逆転条件を上回った. これらから本研究の 2 つめの仮説 (「大ききイメージ」によって生成した身体的イメージは人間の認識を復元で

きる)が部分的に支持される。

しかし、本研究の仮説とは異なり、いくつかのカテゴリのサイズにおいて、逆転条件が「大きさイメージ」条件を上回った。この結果は、先述した単語分散表現に埋め込まれる意味のカテゴリの偏りによってもたらされたと考えられる。

6 結論

本研究では、「大きさ」に関する身体的イメージの生成手法を提案した。提案手法を検証する2つの実験により、単語分散表現から人間の認識と整合する「大きさイメージ」を生成できること、「大きさイメージ」から生成した身体的イメージが人間の認識を復元できることを部分的に確かめた。

本研究の結果の限界は、一部のカテゴリにおいては人間の認識と整合する身体イメージの生成が行えなかったことである。4.1.4節で述べたように、単語分散表現上に埋め込まれる属性は、身体的・物理的なものというより、抽象的なものと考えられる。この性質に対応するためには、今後、カテゴリごとに正規化の方法を検討するなどの改良を行う必要がある。

また、今回は「大きさイメージ」条件との対比として、逆転条件を設定した。今後は、本研究で提案した手法の各段階の妥当性を検討する対比条件を設定した検討も行う必要がある。たとえば、図4における下位のsynsetの組み合わせを用いた「大きさの指標」によって生成した身体イメージを対比条件として設定することもできる。

上記の課題に加えて、将来的には「大きさ」だけでなく、「鋭さ」や「速さ」など様々な指標での身体的イメージ生成に取り組む。これによって、冒頭で述べた、「人間の記号と数量の変換機構を搭載した、人間とのスムーズなインタラクションを実現する高度なヒューマンインタフェース」の発展に貢献できると考える。

謝辞

本研究はJSPS科研費JPJP19H05691, JP20H05560, JP22H04861の助成を受けたものです。

参考文献

- [1] Barbara Tversky. Mind in motion: How action shapes thought. Hachette UK, 2019.
- [2] A Thousand Brains: A New Theory of Intelligence. Basic Books, 2021.
- [3] Steven Pinker. The stuff of thought: Language as a window into human nature. Penguin, 2007.
- [4] Yoshua Bengio, Réjean Ducharme, and Pascal Vincent. A neural probabilistic language model. Advances in neural information processing systems, Vol. 13, , 2000.
- [5] Zellig S Harris. Distributional structure. Word, Vol. 10, No. 2-3, pp. 146–162, 1954.
- [6] Akira Utsumi. Exploring what is encoded in distributional word vectors: A neurobiologically motivated analysis. Cognitive Science, Vol. 44, No. 6, p. e12844, 2020.
- [7] Gabriel Grand, Idan Asher Blank, Francisco Pereira, and Evelina Fedorenko. Semantic projection: recovering human knowledge of multiple, distinct object features from word embeddings. arXiv preprint arXiv:1802.01241, 2018.
- [8] 下川原(佐藤)英理, 関野遥香, 李有てい, 黒田知士, 山口亨. ロボットのジェスチャー表現が行動受諾に与える影響. 知能と情報(日本知能情報ファジィ学会誌), Vol. 34, No. 1, pp. 527–532, 2022.
- [9] 福原佑貴, 上出純平, 中野有紀子. コミュニケーションロボットにおけるジェスチャー表出の有効性. 人工知能学会全国大会論文集 第25回(2011), pp. 2D2OS5b3–2D2OS5b3. 一般社団法人人工知能学会, 2011.
- [10] 鈴木正敏, 松田耕史, 関根聡, 岡崎直観, 乾健太郎. Wikipedia記事に対する拡張固有表現ラベルの多重付与. 言語処理学会第22回年次大会発表論文集, pp. 797–800, 2016.
- [11] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.
- [12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
- [13] Francis Bond, Hitoshi Isahara, Kiyotaka Uchimoto, Takayuki Kuribayashi, and Kyoko Kanzaki. Extending the japanese wordnet. In Proc. 15th Annual Meeting of the Association for Natural Language Processing, pp. 80–83, 2009.