

# 不定自然変換理論と Word2Vec を用いた比喩生成 Metaphor generation using indefinite natural transformation theory and Word2Vec

阿久津 規介<sup>1</sup>, 池田 駿介<sup>1</sup>, 布山 美慕<sup>2</sup>, 西郷 甲矢人<sup>3</sup>, 高橋 達二<sup>1</sup>

Kisuke Akutsu, Syunsuke Ikeda, Tatsuji Takahashi, Miho Fuyama, Hayato Saigo

<sup>1</sup> 東京電機大学, <sup>2</sup> 立命館大学, <sup>3</sup> 長浜バイオ大学

Tokyo Denki University, Ritsumeikan University, Nagahama Institute of Bio-Science and Technology  
tatsujit@mail.dendai.ac.jp

## 概要

不定自然変換理論 (TINT: Theory of Indeterminate Natural Transformation) とは、数学の一分野である圏論に基づく動的な比喩理解モデルである。この論文では、TINT で使用するデータとして分散表現を利用できるか検証するために、3つの手法で喩辞・被喩辞の連想イメージを取得した。1つ目は、先行研究と同様の連想イメージ、2つ目は、喩辞・被喩辞間の連想関係を考慮しない場合の連想イメージ、3つ目は、喩辞・被喩辞間の連想関係を考慮した場合の連想イメージである。つぎに、それぞれの連想イメージを用いてシミュレーションを行った。分散表現データは、喩辞・被喩辞の連想イメージの選択と、イメージ間の連想確率をコサイン類似度で代用するときを使用した。TINT に使用するデータとして分散表現を用いることで、人から喩辞・被喩辞の連想イメージ及びイメージ間の連想確率を取得する必要がなくなり、使用できるイメージの数が増える。シミュレーションの結果、人と同様の連想イメージを用いた場合と喩辞・被喩辞の連想イメージ間の連想確率を考慮した場合では比喩的な対応づけが生成された。また、喩辞・被喩辞間の連想確率がある程度なければモデルが上手く動作しないことがわかった。

キーワード：圏論，比喩，類似，類推

## 1. はじめに

比喩は人間の思考過程の基礎となる認知現象である(平・楠見, 2011)。近年、動的な比喩理解モデルとして不定自然変換理論 (TINT: Theory of Indeterminate Natural Transformation) が提案された(布山・西郷, 2018; 池田他, 2021)。TINT ではイメージの意味をイメージ間の連想関係の総体として定義しており、シミュレーションを行うためには喩辞・被喩辞の連想イメージの選択及び使用する全てのイメージ間の連想確率を実験で取得する必要がある。本論文では、TINT

で使用するデータとして Word2Vec で作成した単語の分散表現を用いることを提案する。具体的には、単語間のコサイン類似度をイメージ間の連想確率として代用した。また、喩辞・被喩辞の連想イメージをコサイン類似度を用いて選択する手法を考案した。単語の分散表現を用いることで、データを実験で取得する必要がなくなり、膨大な量のイメージを用いてシミュレーションを行うことができるようになる。また、新しいイメージを使用する場合でも、追加実験をせずにシミュレーションが可能となる。

## 2. 圏論

圏論とは、対象よりも対象間の関係性を重視した数学分野の1つである(西郷・能美, 2019; Awodey, 2008)。圏論の基本概念は、圏、関手、自然変換である。圏  $C$  とは、対象の集まり  $Ob(C)$ 、すべての対象の対  $A, B$  に対する射の集まり  $C(A, B)$  からなる。圏内の射は合成可能で、すべての対象が恒等射をもち、結合律、恒等射律を満たす必要がある。圏論において、対象とは上記に上げた圏の公理を満たすものなら何でもよい。また、すべての対象が持つ恒等射をその対象自身とみなせるため、対象よりも射が重要となる。関手とは、ある圏の構造(対象, 射, 合成関係)を保ったまま他の圏へと対応づけを行うものである。自然変換とは、圏  $A$  から圏  $B$  への関手  $F$  を同じく圏  $A$  から圏  $B$  への関手  $F'$  へと関手の構造を保ったまま写すものである。この対応付けはすべて圏  $B$  の内部で行われる。圏論では対象を圏、射を関手とする「圏の圏」や、圏  $A$  から圏  $B$  への関手を対象、それらの間の自然変換を射とする「関手圏」を定義することができ、メタレベルの関係を上手く表現できる。

## 3. 比喩

比喩とは、「A は B(のよう)だ」という言語現象のことである。心理学において、喩えられる対象  $A$  を

「主題」、喩える対象  $B$  を「喩辞」と呼ぶ。本論文では先行研究(布山他, 2018)と同様  $A$  を「被喩辞」、 $B$  を「喩辞」と呼ぶ。喩は人間の思考活動の基礎となるものであり、概念や言語理解を解明するための鍵となる重要な認知現象である(平・楠見, 2011)。喩辞と被喩辞を比較するために重要となるのは、喩辞と被喩辞の類似性である。「原子は太陽系のような」という喩では、原子の構造と太陽系の構造が類似している。「笑顔は花のような」という喩では、笑顔と花が共通する「美しい」という属性を持つ。このように、喩辞を被喩辞で喩える場合、2つの間に類似性が見出される。喩理解のモデルにおいて、類似性は表象モデルをどう定義するかによってその表現が異なる。本論文で取り扱うモデル(TINT)は、構造的類似性を比較することで喩をモデル化している。

#### 4. Word2Vec

Word2Vec は 2013 年に Tomas Mikolove が提案した自然言語処理の手法で、文章中の単語を多次元ベクトル(分散表現)に変換して単語の意味を表現しようとするものである(Mikolove et al., 2013)。Word2Vec の設計思想には、「同じ文脈に現れる単語は類似した意味を持つ」という分布仮説がある。Word2vec では skip-gram と CBOW(Continuous Bag-of-Words) の 2つのモデルが提案されている。CBOW では、文章  $S(w_1, w_2, \dots, w_n)$  が与えられたときその  $i$  番目に位置する単語  $w_i$  をその周りの単語  $C = (w_{i-c}, \dots, w_{i-1}, w_{i+1}, \dots, w_{i+c})$  ( $c$ : window size) から予測することでモデルを作る。skip-gram は CBOW を逆転させたモデルである。CBOW における単語ベクトルの取得は、文章中から特定の単語が現れる条件付き確率を最大化することに対応し、skip-gram における単語ベクトルの取得は、対象単語のクラスタリングでの誤りの最小化を意味する。

#### 5. 不定自然変換理論

動的な喩理解モデルとして、不定自然変換理論(TINT)がある。TINT の基本構造は圏論によって定義づけられている。また TINT では、喩の動的過程を説明するため確率過程が導入されている。TINT では、あるイメージの意味は他のイメージとの連想関係の総体と考える。例えば、「蝶」というイメージからは、「美しい」「昆虫」「花」「羽」「きれい」「小さい」といったイメージが連想される。これらのイメージへの連想関係の総体が「蝶」というイメージの意味である。

#### 5.1 TINT における喩理解過程

TINT で使用する概念とそれを用いた喩理解過程を示す。ここで、コスライス圏は圏論における概念でありイメージの圏は TINT における概念である。

**定義** コスライス圏:  $C$  を圏,  $X$  を  $C$  の対象とするとき, コスライス圏  $X \setminus C$  とは, 対象を  $\text{dom}(f) = X$  となる全ての射  $f \in C$ , 射を  $X \setminus C$  の対象  $f: X \rightarrow a$  から  $g: X \rightarrow b$  への,  $h \circ f = g$  を満たすような  $h \in C$  とする圏である。また, 射の合成と恒等射は  $C$  での合成, 恒等射である。

**定義** イメージの圏 対象をイメージ, 射をイメージ間の連想関係とする圏であり, それぞれの対象間の射はたかだか 1 つである(痩せた圏)。

**定義** イメージの圏  $C$  における対象  $A$  の意味は, コスライス圏  $A \setminus C$  である。

まず, 「 $A$  は  $B$  のようだ」という喩により, 射  $f: A \rightarrow B$  が与えられる。この  $f$  により, 自然な関手  $BMF$ (base of metaphor functor) が生じる。その後  $BMF$  よりも妥当な関手  $F$  を探索し,  $BMF$  から  $F$  へと自然変換する。この探索された  $F$  が喩の解釈である。

#### 5.2 圏の不定化

前節における TINT の喩理解過程では、自然変換の探索や喩によるイメージの意味の変化といった動的な活動を表現できていない。TINT では、圏論に確率過程を導入することでこの問題を解決する。

まず潜在圏と顕在圏を作る。

**定義** 潜在圏 全てのイメージ間の連想関係をモデル化したものを顕在圏と呼ぶ。各射  $f_i$  はそれぞれの域から余域への連想確率をもつ。この圏は全結合の構造をもつ。

**定義** 顕在圏 ある時点での励起した射すべてを含む  $C_{exe}$  を顕在圏と呼ぶ。

潜在圏と顕在圏を導入することで、射に連想確率が与えられ、射の励起・緩和過程が追加された。以下に、圏の不定化を導入した TINT の概要を示す。

- イメージ  $A$  の意味は、コスライス圏  $A \setminus C$  と各射  $f_i$  の持つ連想確率  $\mu_i$  の総体である。
- 「 $A$  は  $B$  のようだ」という喩が与えられると顕在圏  $C_{exe}$  に射  $f: A \rightarrow B$  が連想確率  $\mu$  に関わらず励起する。
- $f$  を契機として、関手  $F$  の探索における射の励起・緩和が発生する。励起・緩和のルールを以下

に示す.

0. 基本規則 (basic rule): 励起した二つの射の合成によってできた射は  $\mu$  に関わらず励起する

1. 近接規則 (neighboring rule): 励起した射の余域を域にもつ射は確率  $\mu$  で励起する

2. 分岐規則 (fork rule): 域を共有している射が励起しているとき, その余域の間の射が確立  $\mu$  によって励起する

3. 反分岐規則 (anti-fork rule): 域を共有する2つの射が互いの余域間に射を持たないとき, この2つの射は緩和される

- 以上の過程を経て, 不定であった圏が顕在圏として定まり, 励起した射の族は  $BMF$  からの自然変換をなす. これが不定自然変換である.
- 不定自然変換の過程により,  $\mu$  の値が変化する. これは, 比喩による学習に対応し, 連想確率  $\mu$  が変化することで, イメージの意味が変化する.

以上が TINT の説明である.

## 6. TINT に関する先行研究

前節では TINT の比喩理解過程の概要を説明したが, 先行研究では TINT に関するより詳しい解説がなされている. TINT の理論に関するより詳しい説明は池田他 (2021) が行っている. 連想ネットワークの動的变化については, 横須賀他 (2021) がシミュレーションを行った. TINT における比喩の性質による違いについての研究は, 池田・高橋 (2021) でなされた.

## 7. Word2Vec を用いたシミュレーション

先行研究 (池田・高橋, 2021) によって, 喩辞の慣習性が中程度と高程度の比喩ではある程度人間に近い対応付けが行われることが示された. 本シミュレーションでは, 喩辞の慣習性が高い「蝶は踊り子のようだ」という比喩 (岡他, 2019) についてシミュレーションを行った. これらのシミュレーションは先行研究 (池田他, 2021) に即して行った.

### 7.1 使用データ取得

今回のデータとして, 2022 年 1 月時点で青空文庫にて公開されている作品 14,758 本を使用した<sup>1</sup>.

<sup>1</sup>まとまった近現代の日本語の文章のデータベースであり, かつ比喩表現が文学や随筆で効果的に多用されることも期待し青空文庫を用いた. しかし本論文の以降の箇所で見られるように, 結果的に時代・文化的制約に由来する偏りが強く現れてしまった部分があると思われる. 今後, 潜在圏の設定に用いるデータの選定や, 扱う比喩の設定に関しては, より適切な方法をやり方を検討する

表 1 2022 年 7 月 6 日における分野別作品数

分類	作品数
総記	134
哲学	367
歴史	647
社会科学	803
自然科学	269
技術, 工学	219
産業	46
芸術, 美術	1281
言語	85
文芸	15135
計	18,087

表 1 に 2022 年 7 月時点での青空文庫にて公開されている作品 18,087 本の分類別作品数を示す. ここで, 表のデータは最近のデータであり, シミュレーションに用いたデータよりも母数が多いことに注意する. 次に, MeCab を用いて取得した文章をトークナイズした. また, stop word として slothlib (<https://ja.osdn.net/projects/slothlib/>) にて公開されているデータにひらがな 1 文字を加えたものを使用した. 最後に, トークナイズしたものに gensim (<https://radimrehurek.com/gensim/>) で公開されている CBOW アルゴリズムを用いてモデルを生成した. 生成したモデルには, 合計 109,154 単語が含まれていた.

### 7.2 コサイン類似度の計算

前節のデータを用いてコサイン類似度を計算した. コサイン類似度の計算方式は式 (1) である. シミュレーションではコサイン類似度を確率として扱うために, コサイン類似度が 0 以下の場合は全て 0 とした (式 (2)).

$$\cos(a, b) = \frac{a_1 b_1 + a_2 b_2 + \dots + a_n b_n}{\sqrt{a_1^2 + \dots + a_n^2} \sqrt{b_1^2 + \dots + b_n^2}} \quad (1)$$

$$\begin{cases} 0 & (\text{cossim}(a, b) \leq 0) \\ \text{cossim}(a, b) & (\text{cossim}(a, b) > 0) \end{cases} \quad (2)$$

### 7.3 先行研究と同様の連想イメージの取得

先行研究 [g] では, 「蝶は踊り子のようだ」という比喩についての人の対応付けと TINT のシミュレーション結果を比較している. そのデータと比較するために, 先行研究と同様の連想イメージを選択した. 作成したモデル内に連想イメージとして使用する単語がない場合は, それと類似した意味を持つ単語を連想イメージとして使用した (表 3). 最終的に選択した連想

表 2 先行研究と同様の連想イメージ

喩辞・被喩辞	連想イメージ
蝶	羽, 儂い, 美しい, 空, 女性, 花, 飛翔, 舞い
踊り子	踊り, 女性, スカート, 夜, 音楽, 回転, 揺れ, 舞台

表 3 先行研究と同様の連想イメージにおける変更点

変更前	変更後
舞う	舞い
美しさ	美しい
儂さ	儂い
飛ぶ	飛翔
揺れる	揺れ
回る	回転

イメージを表 2 に示す。また、選択した喩辞・被喩辞の連想イメージ間のコサイン類似度を図 1 に示す。また、先行研究で行われた実験で取得した人の連想確率を図 2 に示す。図 1, 2 を比較すると、[踊り, 舞い(舞う)][舞台, 舞い(舞う)][女性, 美しい(美しさ)]のコサイン類似度は人とコサイン類似度を用いた場合でどちらも高い値となっており、類似した連想確率となっている。人の場合に確率が 0.95 と高い[スカート, 女性]については、コサイン類似度を用いた場合では 0.27 と低なっている。他にも人の連想確率とコサイン類似度との差異が見られる。以上から、コサイン類似度を用いることで、イメージ間連想確率を部分的に捉えてられていることがわかる。

取得した連想イメージ及び連想イメージ間のコサイン類似度を用いてシミュレーションを行った。

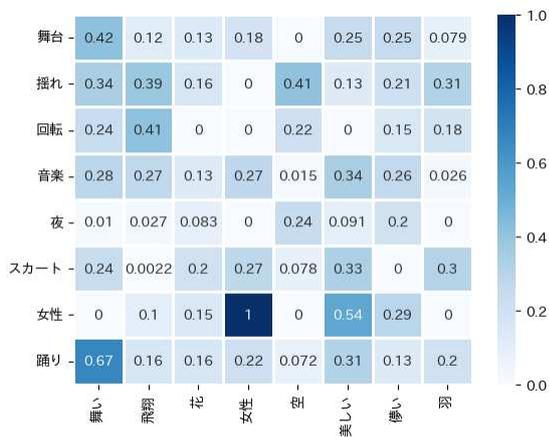


図 1 喩辞・被喩辞の連想イメージ間のコサイン類似度 (先行研究と同様の連想イメージ)

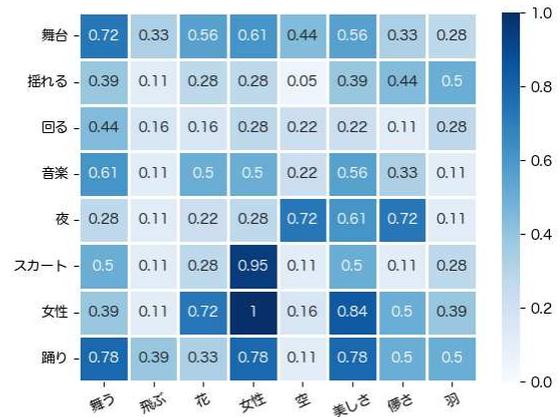


図 2 喩辞・被喩辞の連想イメージ間のコサイン類似度 (人)

### 7.4 喩辞・被喩辞間の連想関係を考慮しない連想イメージの取得

コサイン類似度を用いて喩辞・被喩辞の連想イメージを選択した。この選択手法では、喩辞・被喩辞間の連想確率を考慮していない。まず、作成したモデル内のすべての単語について喩辞とのコサイン類似度を計算した。次に、コサイン類似度の大きい順にソートを行い、上位 8 単語を抽出した。この抽出した単語 8 つを喩辞の連想イメージとした。被喩辞も同様のプロセスを用いて連想イメージを選択した。取得した喩辞・被喩辞の連想イメージを表 4 に示す。また、喩辞・被喩辞の連想イメージ間のコサイン類似度を図 3 に示す。

図 3 を見ると、蝶の連想イメージである「揚羽」「蝶々」「胡蝶」は踊り子の連想イメージとのコサイン類似度がある程度あり、踊り子の連想イメージである「マネキン」「客席」「踊子」は蝶の連想イメージとの間にコサイン類似度がある程度あったが、それ以外のコサイン類似度はほとんど 0 となっている。

表 4 について、蝶の連想イメージを見ると、蝶と同じ概念集合に属するものが多く、蝶から連想するイメージとして偏りが生じている。また、「揚羽」「永生」「胡蝶」など日常生活で使われることが少ないイメージが多く含まれている。

取得した連想イメージ及び連想イメージ間のコサイン類似度を用いてシミュレーションを行った。

表 4 喩辞・被喩辞の連想イメージ間 (喩辞・被喩辞間の連想関係を考慮しない)

喩辞・被喩辞	連想イメージ
蝶	揚羽, 永生, 蝶々, 翅, 桔梗, 蛾, 胡蝶, 蜻蛉
踊り子	踊子, ダンサー, ステージ, レビュー, 客席, マネキン, オペラ, ガール

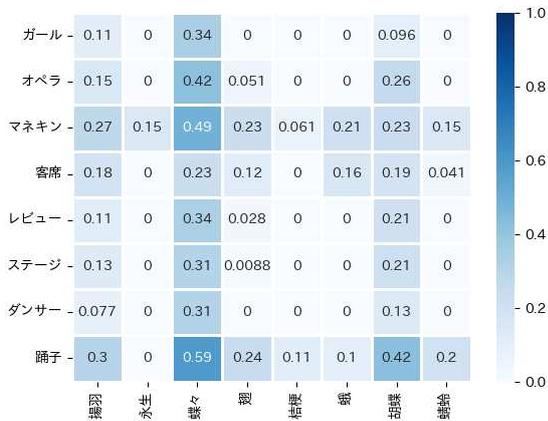


図 3 喩辞・被喩辞の連想イメージ間のコサイン類似度 (喩辞・被喩辞間の連想関係を考慮しない)

### 7.5 喩辞・被喩辞間の連想関係を考慮する連想イメージ

コサイン類似度を用いて喩辞・被喩辞の連想イメージを選択した。前節とは異なり、こちらの選択手法では喩辞・被喩辞の連想イメージ間の連想確率を考慮した。以下に連想イメージ選択のアルゴリズムを示す。

1. 作成したモデル内の全ての単語について、喩辞・被喩辞とのコサイン類似度を計算
2. 喩辞・被喩辞それぞれとのコサイン類似度について、上位 100 単語を抽出
3. 喩辞 (被喩辞) のイメージとして抽出された 100 単語それぞれについて、被喩辞 (喩辞) のイメージとして抽出された単語 100 単語とのコサイン類似度と、喩辞 (被喩辞) とのコサイン類似度の合計値を計算 (式 (3))
4. 合計値の大きい順にソートを行い、上位 8 単語を連想イメージとして採用

表 5 喩辞・被喩辞の連想イメージ間 (喩辞・被喩辞間の連想関係を考慮する)

喩辞・被喩辞	連想イメージ
蝶	紅雀, 胡蝶, 花々, 紋白, 雌蝶, 艶色, 蝶々, 妖艶
踊り子	妖艶, 男役, ガールズ, 緞帳, 天勝, 歌うたい, メーキャップ, 舞妓

$a$  : 被喩辞 (喩辞),  $b$  : 喩辞 (被喩辞)

$a$ -image  $\in A$  ( $a$  とのコサイン類似度上位 100 単語),

$b$ -image  $\in B$  ( $b$  のコサイン類似度上位 100 単語),

weight-sum( $a$ -image) =

$$\sum_B \text{cossim}(a\text{-image}, b\text{-image}) + \text{cossim}(a\text{-image}, a) \quad (3)$$

取得した喩辞・被喩辞の連想イメージを表 5 に示す。また、喩辞・被喩辞の連想イメージ間のコサイン類似度を図 4 に示す。

表 5 について、蝶の連想イメージをみると「花々」「艶色」「妖艶」といった蝶が持つ属性のイメージが選択されていた。また、蝶と踊り子で共通する連想イメージ「妖艶」が選択された (属性的類似性)。蝶の連想イメージのうち「紅雀」「胡蝶」「紋白」「雌蝶」は日常的に用いられる単語ではなく、比喩表現に向かない単語だった。また、踊り子の連想イメージである「緞帳」「天勝」「メーキャップ」も同様に、比喩表現に向かない単語だった。図 4 を見ると、図 3 と比べて全体的に高い値となっており、より人の連想イメージ (図 2) に近づいている。しかし、人の場合ほど連想確率の値に差異はなかった。

取得した連想イメージ及び連想イメージ間のコサイン類似度を用いてシミュレーションを行った。

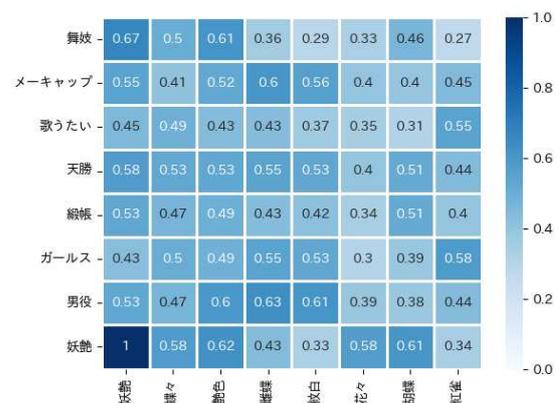


図 4 喩辞・被喩辞の連想イメージ間のコサイン類似度 (喩辞・被喩辞間の連想関係を考慮する)

表 6 対応づけ回数上位 8 組 (先行研究と同様の連想イメージ))

踊り子の連想イメージ	蝶の連想イメージ	対応付回数
女性	女性	609
踊り	舞い	447
揺れ	空	404
舞台	舞い	358
回転	飛翔	311
女性	美しい	302
スカート	羽	287
女性	儂い	235

### 8. 結果

それぞれの手法で選択した連想イメージでのシミュレーションごとに結果をまとめて示す。それぞれの選択手法で選択した連想イメージを用いたシミュレーション結果を、図 5, 6, 7 に示す。図 5, 6, 7 は三角構造の対応づけを時間発展させず、シミュレーションを 1,000 回行ったときに、喩辞の意味を構成するあるイメージが被喩辞の意味を構成するイメージに何回対応づけられたかを示しており、全ての三角構造の対応づけの内、対応づけられた回数が最も多い組が入っている三角構造の対応づけである。また、それぞれのシミュレーションにおいて、対応づけられた回数が多い上位 8 組をそれぞれ表 6, 7, 8 に示す。対応づけ回数は、被喩辞のコスライス圏における全ての三角構造についてシミュレーションを行った結果から、最も対応づけの回数が多いものを代表値として用いている。図 5 を見ると、比喩的な対応づけである、[踊り, 舞い][女性, 美しい][女性, 儂い][踊り, 羽] の対応づけの回数が多く、比喩的ではない [踊り, 空][女性, 空][女性, 羽][女性, 飛翔] の対応づけの回数が低くなった。図 6 を見ると、喩辞・被喩辞の連想イメージ間のコサイン類似度がある程度あった「揚羽」「蝶々」「胡蝶」が多く対応づけられている。また、対応付けられなかった回数 (NA) が全体の 2 割を締めていた。表 7 を見ると、対応づけの回数は最大でも 300 であり、2~3 割しか対応付けがされなかった。これは表 6, 8 それぞれの最も多い対応づけ回数の約半分の値であり他の連想イメージ選択手法と比べて対応づけが行われていないことがわかった。

図 7 を見ると、対応づけがされなかった回数は 5 回であり、殆どの場合対応づけがされた。また、対応づけは [妖艶, 胡蝶][ガールズ, 雌蝶] がそれぞれ全体の 5 割以上を締めており、対応づけられるイメージに偏りが生じていた。表 8 を見ると、表 6 と類似した対応づけ回数であったが、表 8 のほうが全体的に多く対応づけされていた。

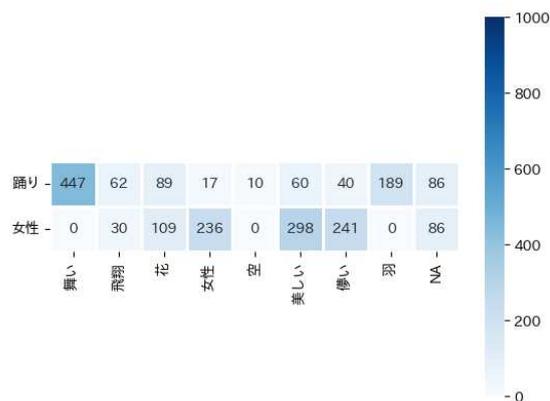


図 5 先行研究と同様の連想イメージを用いたシミュレーション)

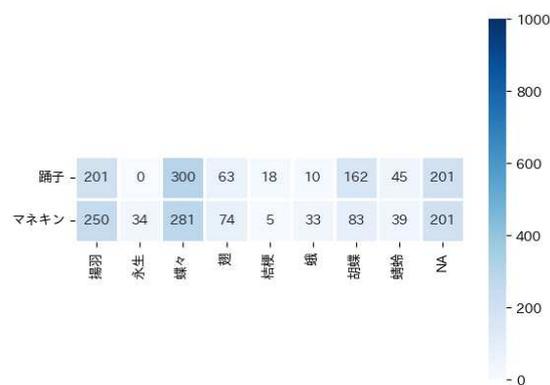


図 6 喩辞・被喩辞間の連想関係を考慮しない場合のシミュレーション)

### 9. 結果の分析と考察

喩辞と被喩辞のコスライス圏を対応付けるために、それぞれの連想イメージを一对一对応させる bijection を作成した。以下に、bijection 選択のアルゴリズムを示す。

0. すべての喩辞・被喩辞の連想イメージの組とその対応づけ回数をリストにまとめる
1. すべての喩辞・被喩辞の連想イメージの組を対応づけ回数の多い順にソートする
2. 最も対応づけ回数が多い組を bijection の結果の一部として取り出す。
3. 取り出された喩辞・被喩辞の連想イメージが含まれる喩辞・被喩辞の連想イメージの組をリストから削除する
4. リストが空でなければ (1) に戻る。リストが空の場合終了。

上記のアルゴリズムを用いて、それぞれのシミュレーション結果から bijection を作成した。作成した bijec-

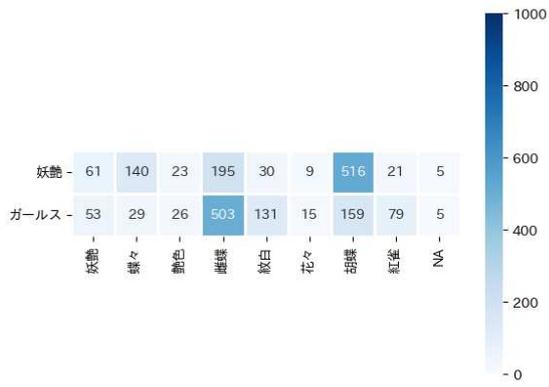


図 7 喩辞・被喩辞間の連想関係を考慮する場合のシミュレーション

表 7 対応づけ回数上位 8 組 (喩辞・被喩辞間の連想関係を考慮しない)

踊り子の連想イメージ	蝶の連想イメージ	対応付回数
ステージ	蝶々	300
オペラ	蝶々	290
マネキン	蝶々	289
マネキン	揚羽	250
ダンサー	蝶々	236
ガール	蝶々	235
レビュー	蝶々	234
ステージ	蝶々	228

tion を、それぞれ表 9, 10, 11 に示す。

表 9 を見ると、[女性, 女性][踊り, 舞い][回転, 飛翔][スカート, 羽]といった比喩的な対応づけが多く見られた。この 4 組は bijection 内の対応づけ回数上位 5 組に含まれており、モデルとして最も妥当な対応づけとして選択されていた。[揺れ, 空]も上位 5 組に入っていたが、比喩的な表現ではなかった。[夜, 花]の対応付けは比喩的な意味が薄く対応付けの回数を見ると 63 回である。これは喩辞・被喩辞のすべての連想イメージが類似した意味を持つのではないことを考えると、どれも妥当な対応づけがされなかったイメージ同士として対応づけされていると考えられる。「蝶にとっての空」と対応されそうな「踊り子にとっての舞台」は対応付けがなされておらず、[揺れ, 空][舞台, 儂い]とどちらも妥当ではない対応付けがされていた。

表 10 は、喩辞・被喩辞の連想イメージ 8 つを全て対応づけすることができなかった。これは表 6, 7, 8 を比較してもわかるように、喩辞・被喩辞の連想イメージ間の連想確率を考慮しない場合には喩辞の連想イメージが「蝶々」「揚羽」といった特定のイメージに偏ったためである。喩辞と被喩辞の連想イメージ間のコサイン類似度は「揚羽, 蝶々, 胡蝶, マネキン, 客席, 踊り子」以外の連想イメージについてはほとんど 0 であったため、喩辞・被喩辞の連想イメージ間の連想

表 8 対応づけ回数上位 8 組 (喩辞・被喩辞間の連想関係を考慮する))

踊り子の連想イメージ	蝶の連想イメージ	対応付回数
妖艶	胡蝶	516
ガールズ	雌蝶	509
歌うたい	蝶々	454
メーキャップ	雌蝶	450
男役	紋白	439
メーキャップ	紋白	438
舞妓	胡蝶	431
男役	雌蝶	419

表 9 喩辞・被喩辞の連想イメージ間の bijection(先行研究と同様の連想イメージ))

踊り子の連想イメージ	蝶の連想イメージ	対応付回数
女性	女性	609
踊り	舞い	447
揺れ	空	404
回転	飛翔	316
スカート	羽	287
音楽	美しい	181
舞台	儂い	166
夜	花	63

確率がある程度なければ対応づけ生じないと考えられる。表 10 で選択された対応づけはどれも意味が読み取れないものであった。

表 11 を考察する。蝶の連想イメージである「胡蝶」は蝶の別名である。故事成語に「胡蝶の夢」があり、これは夢と現実の違いは実ははっきりとしないという意味である。このことを考慮すると、シミュレーションで生成された比喩である「蝶にとっての胡蝶は踊り子にとっての妖艶」は、夢と現実が曖昧な感覚が踊り子を見たときに感じる妖艶さによる現実感の欠落を例えているように捉えられ、比喩が生成されていると考えられる。[ガールズ, 雌蝶]の対応づけはともに「女性」という属性を持つイメージが対応づけされており、比喩理解における属性的類似性を見て取れる。[歌うたい, 蝶々]は、蝶にとっての蝶々が踊り子にとっての歌うたいであり、それぞれ(同じ場所にいる)仲間といった意味で似ていた。これらの組はそれぞれ対応付回数が半分近い値であり、モデルとして最も妥当な対応づけとして選択されていた。

### 10. まとめ

本論文では、TINT で使用するデータとして単語の分散表現を利用できるか検証した。単語の分散表現は喩辞・被喩辞の連想イメージの選択と、イメージ間の連想確率の代わりとしてコサイン類似度を計算する際に使用した。その結果、人と同様の連想イメージを用いた場合と喩辞・被喩辞の連想イメージ間の連想確率を考慮した場合には比喩的な対応づけが生成された。喩辞・被喩辞の連想イメージ間の連想確率を考慮しな

表 10 喩辞・被喩辞の連想イメージ間の bijection(喩辞・被喩辞間の連想確率を考慮しない))

踊り子の連想イメージ	蝶の連想イメージ	対応回数
踊子	蝶々	300
マネキン	揚羽	250
ステージ	胡蝶	145
客席	翅	84

岡 隆之介, 大島 裕明, 楠見 (2019) 孝. 比喩研究のための直喩刺激-解釈セット作成および妥当性の検討. 心理学研究. Vol.90, pp.53-62.  
<https://doi.org/10.4992/jjpsy.90.17236>

表 11 喩辞・被喩辞の連想イメージ間の bijection(喩辞・被喩辞間の連想確率を考慮する))

踊り子の連想イメージ	蝶の連想イメージ	対応回数
妖艶	胡蝶	516
ガールズ	雌蝶	509
歌うたい	蝶々	454
男役	紋白	439
天勝	紅雀	227
メーキャップ	妖艶	138
緞帳	艶色	117
舞妓	花々	33

い場合では対応づけ自体が多く行われなかったため、TINT において喩辞・被喩辞間の連想確率がある程度なければ対応づけが行われないことがわかった。

今後の課題をいくつか挙げる。今回のシミュレーション結果における対応づけが比喩らしいかを筆者が独断で決めていたため、今回のシミュレーションで生成した比喩がどれくらい比喩らしいかを実験する必要がある。また、コサイン類似度とイメージ間の連想確率との関係について認知実験を行い、2つの差異を明らかにする必要がある。

## 文献

- 西郷 甲矢人, 脳美 十三 (2019). 圏論の道案内 技術評論社.
- Steve Awodey (2008). Category Theory second edition. Oxford University Press.
- 平 知宏, 楠見 孝 (2011). 比喩研究の動向と展望. 心理学研究, 82(3), pp.283-299.  
<https://doi.org/10.4992/jjpsy.82.283>
- 布山 美慕, 西郷 甲矢人 (2018). 不定自然変換理論の構築: 圏論を用いた動的な比喩理解の記述. 知識共創, (8), pp.1-11.
- Tomsa Mikolove, Kai Chen, Greg Corrado, Jeffrey Dean (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint, arXiv: 1301.3781.  
<https://doi.org/10.48550/arXiv.1301.3781>
- slothlib <https://ja.osdn.net/projects/slothlib/>
- 池田 駿介, 布山 美慕, 西郷 甲矢人, 高橋 達二 (2021). 不定自然変換理論に基づく比喩理解モデルの計算論的実装の試み. 認知科学, (28), P39-56.  
<https://doi.org/10.11225/cs.2020.065>
- 横須賀 天臣, 池田 駿介, 布山 美慕, 西郷 甲矢人, 高橋 達二 (2021). 不定自然変換理論による連想ネットワークの動的変化. 人工知能学会全国大会, (35).  
[https://doi.org/10.11517/pjsai.JSAI2021.0\\_1H2GS1a03](https://doi.org/10.11517/pjsai.JSAI2021.0_1H2GS1a03)
- 池田 駿介, 高橋 達二 (2021) 不定自然変換理論に基づく比喩理解モデルの計算論的実装と実験的検証. 東京電機大学情報学研究科修士論文 (未公開)
- gensim <https://radimrehurek.com/gensim/>