

# WordNet 収録語彙における意味的距離の最適化表示に向けた制御

金崎航平<sup>\*1</sup>・小野真嗣<sup>\*2</sup>・曾我聡起<sup>\*1</sup>

Email: m2250080@photon.chitose.ac.jp

\*1: 公立千歳科学技術大学 大学院理工学研究科理工学専攻

\*2: 北見工業大学 工学部 地域国際系

◎Key Words

WordNet, 反意語, 意味的距離

## 1. はじめに

本研究は、英語学習者が既習語の一部しか活用できていない現状を改善し、語彙の運用力を向上させることを目指す。語彙の意味的関係を理解することは、語彙知識の深化や文脈理解において重要な要素である。学習者に関連語を効果的にグルーピングして学習させるため、文字情報に加え視覚的な図示情報を提供する。

筆者ら(2025)<sup>(1)(2)</sup>の先行研究では、英語語彙を意味関係に基づいて体系化した語彙データベースである WordNet<sup>(3)</sup>の上位語(hyponym)・下位語(hyponym)・同義語(synonym)の関係を Universal Sentence Encoder(USE)<sup>(4)</sup>を用いてコサイン類似度を算出し、語彙間の意味的距離を可視化するアプリを開発した。その結果、学習者が語彙関係を視覚的に理解する有効性が確認された。

さらに、WordNet 全体を対象とした語彙関係の分析を行った結果、反意語(antonym)は意味的に対立関係にあるが、高いコサイン類似度が算出される問題が確認された。この結果、可視化図上でも反意語が近くに配置され、学習者に意味的関係の誤解を生じさせる可能性が示唆される。

そこで本研究では、Counter-Fitting<sup>(5)</sup>を用いて反意語間の意味的距離を制御し、人間の意味理解に近い意味空間を構築する。さらに、SimLex-999<sup>(6)</sup>を用いて意味空間の妥当性を評価するとともに、二次元および三次元可視化による語彙学習支援の可能性を検討する。

## 2. 本研究の目的

本研究では、反意語間の意味的距離を制御した意味空間を構築し、その有効性を評価することを目的とする。具体的には以下の3点を目的とする。

第一に、反意語が定義文中の共通語彙の影響により高い類似度を示す問題に対し、Counter-Fitting を用いて反意語間の意味的距離を適切に制御することである。

第二に、構築した意味空間が人間の意味的類似度判断とどの程度一致するかについて、SimLex-999 を用いて定量的に評価し、その妥当性を検証することである。

第三に、構築した意味空間を二次元および三次元で可視化し、可視化手法の違いが語彙間の意味理解に与える影響を検討することである。

## 3. 研究方法

本研究では、Counter-Fitting を用いて反意語間の意味的距離を制御し、SimLex-999 による評価を行った。また、構築した意味空間を二次元および三次元で可視化し、その有効性を検証した。次節以降で、各手法および評価方法の詳細について記述する。

### 3.1 コサイン類似度の算出方法

英語語彙間の意味的距離を算出するため、WordNet に収録された各語彙の定義文を利用した。各語彙がもつ複数の定義文をピリオドで連結し、一続きの文章として扱

い、文単位の意味情報を高精度にベクトル化できる USE によってベクトル化を行った。USE は文全体の意味を保持した分散表現を生成する文埋め込みモデルであり、WordNet の定義文のような長文テキストから語彙の意味を表現する手法として適しているため、本研究において採用した。これにより、1つの単語に含まれるすべての語義情報を統合した上で、語彙全体を1つの意味ベクトルとして扱うことが可能となる。この手法によって得られたベクトル数は、148,731個であり、今回抽出した語彙数に対応している。

語彙間の意味的類似度は、ベクトル間のコサイン類似度によって算出した。コサイン類似度は-1から1の範囲をとり、値が大きいほど意味的に近いことを示す。

一例として、“absent”の定義文を以下に示す。

”go away or leave. not being in a specified place. nonexistent. lost in thought; showing preoccupation.”

### 3.2 Counter-Fitting による反意語制御

先行研究の結果、反意語は意味的に対立関係にあるが、定義文には共通する文脈や語彙が多く含まれるため、必ずしも低い類似度を示すとは限らず、高いコサイン類似度が算出されるという問題が確認された。

そこで本研究では、Counter-Fitting を適用し、語彙ベクトル空間の再構築を行った。Counter-Fitting は、同義語を近づけ、反意語を遠ざけるようにベクトルを調整する手法である。

本手法は、既存の単語埋め込みを再学習するのではなく、後処理として言語学的制約を付与する手法のため、元の意味空間を大きく壊さずにベクトルを調整できるという利点を持つ。

本研究では、WordNet から取得した反意語および同義語の関係を制約条件として利用し、反意語間の意味的距離を拡大することで、人間の意味理解に近い意味空間の構築を試みた。

また、Counter-Fitting では、反意語を引き離す項 (AR)、同義語を引き寄せる項 (SA)、および元の意味空間を保持する項 (VSP) からなる目的関数を用いてベクトルを最適化する。主な制御パラメータとして、反意語間に要求する最低距離を表す  $\delta$ 、同義語間の目標距離を表す  $\gamma$ 、反意語を引き離す強さを表す  $k_1$ 、同義語を引き寄せる強さを表す  $k_2$ 、元の意味空間を保持する強さを表す  $k_3$  を用いる。また、最適化には確率的勾配降下法 (SGD) を用い、SGD ステップ数をベクトル更新回数として設定した。目的関数は以下のように定義され、これを最小化することで語彙ベクトルの最適化を行う。

$$L = k_1 AR + k_2 SA + k_3 VSP$$

### 3.3 SimLex-999 による妥当性評価

Counter-Fitting による反意語制御の有効性を検証するため、SimLex-999 を用いた評価を行った。SimLex-999 は、

人間が評価した単語対の意味的類似度を収録したデータセットであり、単なる関連性ではなく意味的類似性を評価対象としている。

本研究では、Counter-Fitting 適用前後の語彙ベクトルについて、各単語対のコサイン類似度を算出し、SimLex-999 の評価値との比較を行った。評価指標として、Pearson 相関係数および Spearman 順位相関係数を用いた。

また、Counter-Fitting のパラメータは、SimLex-999 との Spearman 順位相関係数が最大となるよう調整した。各パラメータを複数の条件で変更しながら評価を行い、最も高い評価値を示した組み合わせを採用した。

### 3.4 可視化システムの構築

Counter-Fitting によって構築した意味空間を学習者が直感的に理解しやすくするため、可視化システムを構築した。

可視化対象には中心語とその関連語である、上位語、下位語、同義語、反意語を用い、語彙間の距離が意味的距離を反映するよう配置した。距離が近い語ほど意味的に近く、距離が遠い語ほど意味的に離れていることを表す。

本研究では、二次元可視化および三次元可視化の2種類を実装した。二次元可視化では平面上に語彙を配置し、全体構造を把握しやすくした。一方、三次元可視化では奥行きを加えることで、語彙群の空間的な広がりやクラス構造を表現できるようにした。

システムは FileMaker の Web ビューア上で動作するように構築した。二次元可視化には Canvas API を用い、三次元可視化には Three.js を用いて描画を行った。また、JavaScript によって描画処理および操作機能を実装し、利用者はドラッグ操作による視点移動や回転、ピンチ操作による拡大・縮小を行うことが可能であり、スマートフォンで操作可能な学習支援アプリとして構築した。

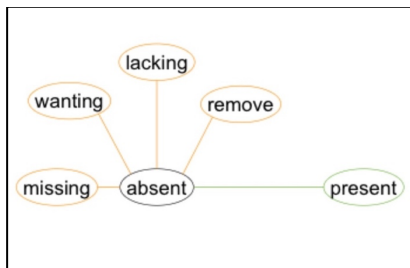


図1 語彙間距離可視化の例(二次元, "absent")

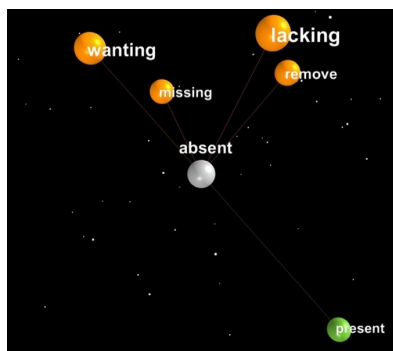


図2 語彙間距離可視化の例(三次元, "absent")

### 3.5 予備実験

構築した可視化システムの有効性を検証するため、予備実験を実施した。本実験では、二次元可視化と三次元可視化を比較し、可視化手法の違いが語彙間の意味理解に与える影響を調査した。

実験で使用する語彙は、CEFR-J Wordlist<sup>7)</sup>を参考に選定した。CEFR-J Wordlist は、日本人英語学習者を対象とし

て語彙を CEFR レベル別(A1~B2)に分類した語彙リストである。本研究では、被験者の語彙知識の影響を考慮し、A1 から B2 レベルの語彙を対象とした。また、WordNet において反意語および同義語の関係が定義されている語を対象とし、可視化図上で語彙間の意味関係を表現できる語を選定した。さらに、各タスク間で語彙の難易度に大きな偏りが生じないように、CEFR レベルを考慮して語彙を選定した。

被験者は公立千歳科学技術大学の学生 6 名とした。実験開始前に操作方法の説明および操作練習を実施した後、本実験を行った。被験者は二次元および三次元可視化の両条件でタスクを実施した。順序効果を抑制するため、被験者を「二次元可視化→三次元可視化」と「三次元可視化→二次元可視化」の2群に分けた。

なお、本予備実験では語彙関係に基づく色分けは行わず、距離情報のみに基づいて可視化図を提示した。これにより、視覚的手がかりを最小限にした条件下で、語彙間距離そのものが意味理解に与える影響を検証した。

実験では、以下の3種類のタスクを実施した。

タスク1: 近い語・遠い語選択

中心語から視覚的に最も近い単語と最も遠い単語を選択するタスクであり、可視化図における距離情報を正しく知覚できるかを評価した。

タスク2: 同義語選択

中心語と意味的に近い語を全て選択するタスクであり、同義語を他の語群と区別して認識できるかを評価した。

タスク3: 反意語選択

中心語と意味的に反対の語を全て選択するタスクであり、反意語を他の語群と区別して認識できるかを評価した。

評価指標として、各タスクの正答率および回答時間を記録した。また、各可視化条件終了後に主観評価アンケートを実施した。アンケートでは、可視化図の見やすさ、意味関係の理解しやすさ、近い意味を持つ単語群の認識しやすさ、反対の意味を持つ単語群の認識しやすさ、単語間距離の理解しやすさ、操作性、学習意欲、有用性および疲労感について5段階評価で回答を求めた。

加えて、タスク2およびタスク3で使用した単語について、被験者が各単語の意味をどの程度理解していたかを確認するため、語彙既知度調査を実施した。調査では、各単語に対して「1: 知らなかった」「2: 見たことはあるが意味はわからなかった」「3: なんとなく意味はわかった」「4: 意味を説明できる」の4段階で回答を求めた。この調査により、タスク成績に対する語彙知識の影響についても検討した。

## 4. Counter-Fitting による反意語制御の評価

本章では、SimLex-999 を用いて Counter-Fitting のパラメータ設定を検討した結果を示す。さらに、選定したパラメータを適用した際のコサイン類似度の変化および SimLex-999 との相関を分析し、反意語制御の有効性について考察する。

### 4.1 Counter-Fitting パラメータの検討

制御後のコサイン類似度と SimLex-999 との Spearman 順位相関係数が最大となるパラメータの組み合わせを探索した結果を表1に示す。

表1より、反意語間に要求する最低距離を表す  $\delta$  および反意語を引き離す強さを表す  $k_1$  は、反意語のコサイン類似度に大きな影響を与えることが確認された。一方、同義語の目標距離を表す  $\gamma$  および同義語を引き寄せる強さを表す  $k_2$  の影響は比較的小さかった。また、元の意味空

間を保持する強さを表す  $k_3$  は、全体の類似度分布に影響を与えることが確認された。

表 2 に、Counter-Fitting 適用前後におけるコサイン類似度の変化を示す。表 2 より反意語の平均コサイン類似度は 0.4546 から -0.6781 へと大きく低下したことが示された。また、中央値についても 0.4500 から -0.7120 へ低下しており、反意語間の意味的距離が明確に拡大されたことが確認できる。

一方、上位語 / 下位語の平均コサイン類似度は 0.3175 から 0.2590 へとわずかに低下し、同義語についても 0.5115 から 0.4834 へとわずかな変化に留まった。この結果から、Counter-Fitting は主として反意語関係に強く作用し、他の語彙関係への影響は比較的限定的であることが示唆される。さらに、反意語のコサイン類似度の差分について Wilcoxon 符号順位検定を行った結果、制御前後で統計的に有意な差が確認された ( $p < 0.0001$ )。このことから、Counter-Fitting による変化は統計的にも有意であることが示された。

表 3 に SimLex-999 との関連評価について示す。表 3 より Counter-Fitting 適用前は Pearson 相関係数が 0.3704, Spearman 順位相関係数が 0.3621 であったのに対し、適用後はそれぞれ 0.4036, 0.4287 へと向上した。特に Spearman 順位相関係数は約 18% の向上が確認された。

表 1 制御パラメータ設定

| パラメータ    | 値    | 説明            |
|----------|------|---------------|
| $\delta$ | 1.4  | 反意語に要求する最低距離  |
| $\gamma$ | 0.0  | 同義語の目標距離      |
| $k_1$    | 5.0  | 反意語を引き離す強さ    |
| $k_2$    | 0.1  | 同義語を引き寄せる強さ   |
| $k_3$    | 0.06 | 元の意味空間を保持する強さ |
| SGD      | 20   | ベクトル更新回数      |

表 2 Counter-Fitting 適用前後のコサイン類似度変化

| 関係種別      | 条件  | 平均      | 中央値     | 標準偏差   |
|-----------|-----|---------|---------|--------|
| 反意語       | 制御前 | 0.4546  | 0.4500  | 0.1691 |
|           | 制御後 | -0.6781 | -0.7120 | 0.2406 |
| 同義語       | 制御前 | 0.5115  | 0.4828  | 0.2177 |
|           | 制御後 | 0.4834  | 0.4482  | 0.3162 |
| 上位語 / 下位語 | 制御前 | 0.3175  | 0.3104  | 0.1366 |
|           | 制御後 | 0.2590  | 0.2606  | 0.1995 |

表 3 SimLex-999 に基づく関連評価

| 条件  | Pearson 相関係数 | Spearman 順位相関係数 |
|-----|--------------|-----------------|
| 制御前 | 0.3704       | 0.3621          |
| 制御後 | 0.4036       | 0.4287          |

## 4.2 考察

Counter-Fitting を用いて反意語間の意味的距離を制御し、語彙ベクトル空間の改善を試みた結果、反意語の平均コサイン類似度は大きく低下し、反意語間の意味的距離が適切に拡大されたことが確認された。この理由として、WordNet の定義文には反意語間で共通語彙が多く含まれ、高い類似度が生じやすい構造的問題に対して、Counter-Fitting は共起による類似度の歪みを制約で補正するため、反意語に強く作用したと考えられる。

一方で、上位語・下位語および同義語においてはコサイン類似度の変化は限定的であり、本手法は反意語関係に対して重点的に作用することが示された。これは、語彙間の意味的距離や構造を維持しながら、一部の関係のみを選択的に調整できるという点で、語彙の構造的理解を支援することが示唆される。

また、SimLex-999 との Spearman 相関係数も向上してお

り、人間の意味的類似度判断との一致度が改善されたことが示された。この改善は、USE による定義文ベースの埋め込みが共起語に依存する構造を持つため、反意語間の距離が適切に拡大されたことにより、「純粋な意味類似度」を評価する SimLex-999 との整合性が高まったためと考えられる。これにより、人間の意味的類似度判断に近づいたことを示され、意味空間が統計的共起構造から認知的意味構造へと再配置された可能性を示唆している。

さらに、本研究では元の意味空間を可能な限り保持することを目的として、反意語に対してのみ制御後のコサイン類似度を利用した。この設計により、既存の上位語・下位語・同義語の関係性を維持しながら、反意語のみを選択的に再配置することが可能となった。

以上より、Counter-Fitting は語彙可視化における反意語制御手法として有効であり、人間の意味理解に近い語彙意味空間の構築に寄与することが示された。

## 5. 予備実験結果

本章では、二次元可視化および三次元可視化を用いた予備実験の結果について述べる。

### 5.1 タスク結果

各タスクにおける正答率および回答時間を表 4 に示す。

タスク 1 では、二次元・三次元いずれの条件においても正答率は 100% であり、天井効果が確認された。このため、本タスクでは可視化条件による差異は確認されなかった。回答時間については、二次元条件が  $21.62 \pm 7.32$  秒、三次元条件が  $25.55 \pm 2.92$  秒であり、二次元条件の方がやや短い傾向が見られた。

タスク 2 では、二次元条件の正答率が 67%、三次元条件が 100% であり、三次元可視化において正答率の向上傾向が見られた。これは、語彙の意味的クラスタ構造が三次元空間においてより把握しやすかった可能性がある。回答時間については、二次元条件が  $27.64 \pm 17.00$  秒、三次元条件が  $22.25 \pm 9.13$  秒であり、三次元条件の方が短い傾向が見られた。しかし、二次元条件では標準偏差が 17.00 秒と大きく、被験者間のばらつきが大きかったことから、回答時間の差については今後、被験者数を増やした検証が必要である。

タスク 3 では、二次元・三次元いずれの条件においても 83% の正答率であり、可視化手法による明確な差は見られなかった。回答時間については、二次元条件が  $14.77 \pm 3.00$  秒、三次元条件が  $20.93 \pm 5.25$  秒であり、二次元条件の方が短い傾向が見られた。

以上より、タスク 1 では天井効果により条件間の差は確認されなかった一方で、タスク 2 では三次元可視化が正答率で改善傾向を示した。またタスク 3 では正答率に差は見られないものの、回答時間には条件による違いが見られた。これらの結果から、可視化手法はタスクの種類によって異なる影響を与える可能性が示唆される。

表 4 各タスクにおける正答率および回答時間

| タスク | 条件  | 正答率(%) | 回答時間(s)           |
|-----|-----|--------|-------------------|
| 1   | 二次元 | 100    | $21.62 \pm 7.32$  |
|     | 三次元 | 100    | $25.55 \pm 2.92$  |
| 2   | 二次元 | 67     | $27.64 \pm 17.00$ |
|     | 三次元 | 100    | $22.25 \pm 9.13$  |
| 3   | 二次元 | 83     | $14.77 \pm 3.00$  |
|     | 三次元 | 83     | $20.93 \pm 5.25$  |

## 5.2 主観評価

主観評価の結果を表5に示す。表5より、全体的な傾向として三次元可視化は「見やすさ」「近い意味のまとまり」「反対の意味のまとまり」「操作性」「学習意欲」「有用性」において二次元可視化を上回る評価を示した。特に「学習意欲」は三次元条件で平均5.00と最も高い値を示しており、学習支援ツールとしての有効性が強く示唆される。また「操作性」においても三次元条件の評価が高く、インタラクティブな操作性が学習体験の向上に寄与した可能性がある。

一方で、「意味理解」および「距離理解」については二次元可視化の方が高い評価を示した。これは、平面上での配置の方が語彙間の距離関係を直感的に把握しやすい可能性を示している。

さらに、「疲労感」については両条件とも低い値を示しており、操作負荷は小さいことが確認された。なお、疲労感は値が低いほど負担が少ないことを意味する。

以上より、三次元可視化は学習意欲や操作性の向上に寄与する一方で、二次元可視化は距離および意味関係の直感的理解に優れていることが示された。したがって、両者はそれぞれ異なる認知的利点を持つ補完的な可視化手法であると考えられる。

表5 二次元および三次元可視化に対する主観評価

| 質問項目       | 二次元       | 三次元       |
|------------|-----------|-----------|
| 見やすさ       | 4.17±0.37 | 4.67±0.47 |
| 意味関係理解     | 4.83±0.37 | 4.17±1.21 |
| 近い意味のまとまり  | 4.17±1.07 | 4.67±0.75 |
| 反対の意味のまとまり | 3.83±1.34 | 4.33±0.94 |
| 距離理解       | 4.50±0.76 | 4.17±0.90 |
| 操作性        | 3.17±0.90 | 4.33±0.75 |
| 学習意欲       | 4.50±0.50 | 5.00±0.00 |
| 有用性        | 4.67±0.47 | 4.83±0.37 |
| 疲労感        | 1.83±1.07 | 1.67±0.75 |

## 5.3 語彙既知度調査

タスク2およびタスク3で使用した単語について、各条件での既知度の平均値および標準偏差を表6に示す。表6より、A1およびA2レベルの語彙では高い既知度が確認される一方で、B1およびB2レベルでは既知度の低下が見られ、特にB1およびB2において標準偏差が増加していることから、語彙知識の個人差が大きくなる傾向が示された。

全語彙を対象とした既知度の平均値は3.35±0.31であり、被験者は対象語彙に対して中程度以上の理解を有していることが確認された。また標準偏差は0.31と小さく、全体として極端な未知語は少ないことが示された。

また、正答語の平均既知度は3.35±0.90、誤答語の平均既知度は3.17±1.21であり、わずかな差が確認された。さらに誤答語のCEFR分布を分析したところ、B1およびB2レベルの語彙が誤答全体の67%を占めている一方で、A1およびA2レベルの基礎的語彙も誤答に含まれていた。

表6 CEFR レベル別語彙既知度および正誤別語彙既知度

| 条件  | 既知度       |
|-----|-----------|
| A1  | 3.78±0.42 |
| A2  | 3.70±0.46 |
| B1  | 3.17±1.11 |
| B2  | 2.71±0.98 |
| 全体  | 3.35±0.31 |
| 正答語 | 3.35±0.90 |
| 誤答語 | 3.17±1.21 |

以上より、語彙難易度の上昇に伴い既知度の低下およびばらつきの増加が確認された。また、タスク成績との比較から、語彙既知度は一定の影響を有するものの、可視化による意味構造理解など他の要因もタスク遂行に関与している可能性が示された。

## 6. おわりに

本研究では、Counter-Fitting を適用することにより反意語間の意味的距離を制御し、人間の意味理解に近い意味空間の構築を試みた。さらに、構築した意味空間を二次元および三次元で可視化し、語彙間の空間構造の理解に対する可視化手法の影響を検討した。

Counter-Fitting により反意語間のコサイン類似度は大幅に低下し、意味空間の構造が適切に調整されたことが確認された。また、SimLex-999 との相関係数の向上から、意味空間の妥当性が改善されたことが示された。

予備実験の結果、三次元可視化は操作性および学習意欲の向上に寄与する一方で、二次元可視化は語彙間距離および意味構造の把握において優位性を示す傾向が見られた。語彙既知度の分析では、CEFR レベルの上昇に伴い既知度の低下およびばらつきの増加が確認され、特にB1、B2レベルにおいて個人差が大きいたことが示された。また、タスク結果に語彙既知度は一定の影響を持つものの、それのみでは結果を十分に説明できず、可視化による構造理解など他の要因の影響が示唆された。

以上より、本研究で提案した語彙可視化手法は、語彙間の意味理解を支援する有効な手法である可能性が示された。一方で、本研究は被験者数が6名と少ないため、統計的検定の解釈には慎重を要する。本予備実験は探索的研究として語彙可視化手法の初期的傾向を把握することを目的としており、一定の妥当性を有すると考えられる。加えて、タスク2における結果は示唆的ではあるが、一般化には追加実験による検証が必要である。

今後は、本予備実験の結果を踏まえ、被験者数を拡大した本実験を実施し、統計的有意性の検証および可視化手法の効果をより厳密に評価する予定である。

## 謝辞

本研究はJSPS 科研費 JP22K02825 の助成を受けたものです。

## 参考文献

- (1) 金崎航平, 小野真嗣, 曾我聡起: 「英語語彙間の語義繋がり可視化アプリ開発-学習者の語彙記憶保持を求めて-」『2025 PC カンファレンス論文集』, pp.126-129(2025)
- (2) 金崎航平, 小野真嗣, 曾我聡起: 「WordNet におけるコサイン類似度を用いた語彙間の分析と学習支援アプリへの応用」『2025 PC カンファレンス北海道論文集』, pp.20-23(2025)
- (3) Miller, G. A.: “WordNet: A Lexical Database for English”. Communications of the ACM, Vol.38, No.11, pp.39-41, (1995)
- (4) TensorFlow 「@tensorflow-models/universal-sentence-encoder」<https://www.jsdelivr.com/package/npm/@tensorflow-models/universal-sentence-encoder>(最終閲覧日 2026 年 6 月 27 日)
- (5) Mrkšić, N., O'Seaghdha, D., Thomson, B., Gašić, M., Rojas-Barahona, L., Su, P.-H., Vandyke, D., Wen, T.-H., and Young, S.: “Counter-fitting Word Vectors to Linguistic Constraints”, Proceedings of NAACL-HLT, pp. 142-148, (2016).
- (6) Hill, F., Reichart, R., and Korhonen, A.: “SimLex-999: Evaluating Semantic Models with (Genuine) Similarity Estimation”, Computational Linguistics, Vol.41, No.4, pp.665-695, (2015).
- (7) CEFR-J Project: “CEFR-J Wordlist and Related Resources”, CEFR-J Project, Tokyo University of Foreign Studies, <https://www.cefr-j.org/research.html>(最終閲覧日 2026 年 6 月 27 日)