

単語の意味空間を心的操作する = 射影

Projection matrices can explain mental operations for semantic space

浅川 伸一[†] and 近藤 公久[‡]

Shin Asakawa and Tadahisa Kondo

東京女子大学[†] and 工学院大学[‡]

Tokyo Women's Christian University[†] and Kogakuin University[‡]

asakawa@ieee.org[†] and tkondo@cc.kogakuin.ac.jp[‡]

概要

文脈効果を心的操作として捉えるモデルを提案した。我々は、意味空間のうち部分空間を文脈とし、部分空間への射影変換によって文脈効果を扱うことができると考えた。このような射影変換を用いることの利点は、階層構造で表現された意味概念では、詳細な記述、操作、検索過程などと直接関連させることが困難な心的操作を表現可能なことである。射影変換が文脈効果を表す心的操作の一部と見なすことで推論や検索過程が記述可能と考えられる。そこで本発表においては、日本語 wikipedia を用いて訓練した単語のベクトル表現集合を意味空間とし、これらに対して射影変換によって文脈効果が得られるかを検討した結果を報告する。本手法は、深層学習との親和性も高く、今後発展が期待できる。

キーワード：word2vec, 単語埋め込み表現, 意味空間, 音韻表現

1. はじめに

文脈効果を心的操作として捉えるモデルを提案する。文脈効果とは、言語理解における事前情報の影響によって解釈が異なることに代表されるが、単語命名課題におけるプライミング効果や、文章の多義性・曖昧性の解消、さらには、視覚における恒常性、錯視、だまし絵、視覚探索における制約など、認知的処理における時間的前後に限らない、状況などを文脈ととらえることが可能である。

自然言語処理においては、単語埋め込みモデル、または、ベクトル空間モデルと総称されるモデル (word2vec:[8]; GloVe:[9]; ELMo:[10]; BERT:[1]) が存在する。すでに、翻訳、分類、要約などの種々の言語課題 (GLUE:[12]) で性能向上が認められている。これは、単語や文を多次元ベクトルとして意味空間を表現していると捉えることができる。我々は、意味空間のうち部分空間を文脈とし、部分空間への射影変換によって文脈効果を扱うことができると考えた。

語彙を構成するベクトルからなる行列を \mathbf{X} で表す。 \mathbf{X} が張る空間 $P(\mathbf{X})$ に各単語を射影し、射影された部分空間内で各語間の距離を測れば所与の目的が近似的に達成される。このような射影変換を用いることの利点は、階層構造で表現された意味概念では、詳細な記述、操作、検索過程などと直接関連させることが困難な心的操作を表現可能なことである。

例えば、階層構造による動物概念モデルでは、トラがネコ科であって哺乳類の下位概念であることは明示的に表現できる一方で、生息地が主としてアジアの密林地帯であるという地理的情報は付加的情報として取り扱わざるを得ない。一方、ベクトル空間モデルでは、多次元ベクトル空間内に、生息地の情報も埋め込まれて表現されている。このため、適切な演算を定義すれば文脈の影響を加味した結果を明示的に得ることが可能となる。

このような演算は、意味情報と音韻情報とを併せ持つオノマトペのような単語表象を扱う際にも有効であると考えられる。この射影変換が文脈効果を表す心的操作の一部と見なすことで推論や検索過程が記述可能と考えられる。

オノマトペを構成する単語群を持ちいたオノマトペ空間に単語を射影することで、単語に近いオノマトペを選択可能であろう。オノマトペ自動生成器についての従来手法では、印象評定に基づく手法が提案されている [14]。意味空間を操作する射影に基づくオノマトペ生成は、心理実験によって得られた印象評定に基づく結果と異なるものとなる可能性がある。

そこで本発表においては、日本語 wikipedia を用いて訓練した単語のベクトル表現集合を意味空間とし、これらに対して射影変換によって文脈効果が得られるかを検討した結果を報告する。本手法は、深層学習との親和性も高く、今後発展が期待できる。

本稿では、上述の単語埋め込みモデルを用いて、課題へ適用可能性を検討した。機械学習の分野では、一般に高次元空間を仮定した空間表象モデルでは、ハブ

(hubness) が存在し解釈が困難になることが知られている [11] ハブネスとは、ある点がデータセット内の他の点の k 個の最近傍に出現する回数の分布に影響を与えることを言う。一般的に用いられるデータでは、このハブの分布は次元の増加に伴って歪んだものとなる。このことから、ベクトル空間モデルの解釈には困難が伴うことが知られている。ハブの存在を解消する手法が提案されてきたが、上記 Mikolov らのベクトルの加減算を用いた推論は、ハブを回避する最も単純な提案の一つであるとみなすことができるだろう。

1.1 単語埋め込みモデル

Mikolov らのベクトル空間モデルは、'king' - 'men' + 'women' = 'queen' のようにベクトルの加減算として類推、推論課題を定義した [8]。単語埋め込みモデルでは、意味のみならず、統語についての推論も可能である。図 1 左では、意味が、同図右では統語情報、この場合複単語の単数-複数-の関係を表している [8]。加えて、図 2 では、各国の国名と対応する首都名との関係を主成分分析して布置したものである。同図では、横軸左に国名、右に首都名が対応する首都名が布置されている。興味深いことに縦軸は、上から下にはほぼユーラシア大陸を東から西に並んでいる。主成分分析は、各項目に対応するベクトルを集めた行列を X とすれば、主成分分析の目的関数は以下で与えられる: $\max_w \{w^T X^T X w - \lambda (w^T w - 1)\}$ 。ここで λ は Lagrange 乗数である。主成分分析は w によって張られる空間への射影行列を求めることに相当する。従って、適切な射影行列を定義できれば、国名と首都との関係と、各国の地理的關係をも表象されていると考えることが可能である。

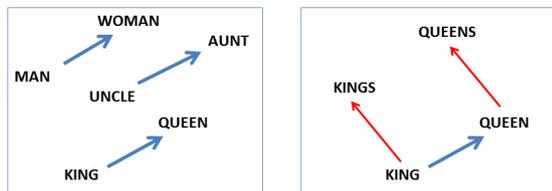


図 1: Word2vec による推論の例: 左は意味。右は文法情報とみなすことができる。 [8] Fig. 2 より

ベクトルの演算を拡張することにより、その他の推論も可能になるように思われる。具体的には、'woman' と - 'men' との両ベクトルが張る部分空間へ、'king' ベクトルを射影した場合、最近傍ベクトルが 'queen' と解釈できる。

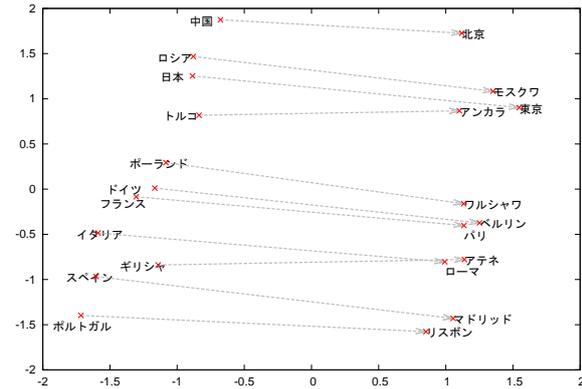


図 2: [7] Fig. 2 を改変。横軸は国とその首都。縦軸はユーラシア大陸を西から東へ大まかな地理情報に対応する

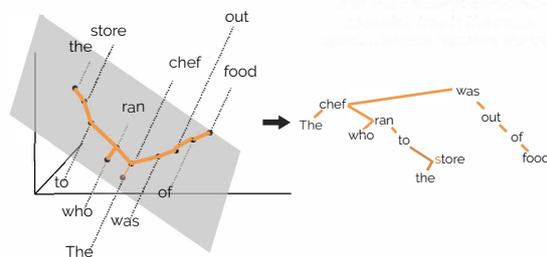


図 3: [3] より。構文解析木のルートからのノード数に対応した距離への変換例。

1.2 線形射影空間

実数ベクトル $X \in \mathbb{R}^{n,m}$ が張る空間への射影行列は次式で定義される:

$$P(X) = X (X^T X)^{-1} X^T \quad (1)$$

本稿では、単語埋め込みモデルで得られる単語ベクトルを用いて意味空間を用いて心的操作を説明することを試みた。

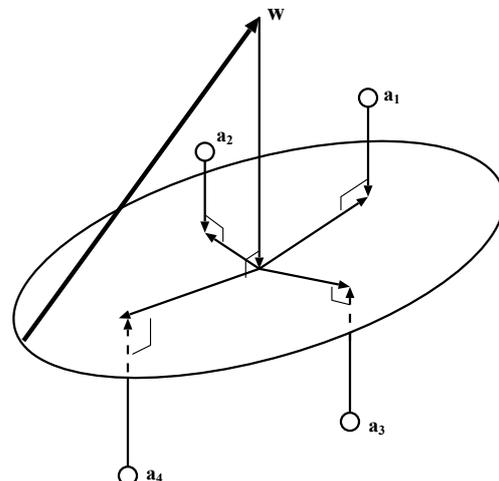


図 4: 射影の概念図

X と Y との類似度を $\text{Sim}(X, Y)$ とする。ベクトル空間内の線形変換 $A = B^T B$ とすれば、射影行列 $B \in \mathbb{R}^{k \times n}$ を用いることにおいて、次式のような距離行列の変換を定義できる:

$$\text{Sim}(X, Y)_B = (B(X - Y))^T (B(X - Y)), \quad (2)$$

変換行列 $\mathbb{R}^{k \times n}$ を学習するには、最小自乗誤差基準を用いるのであれば、次式で求めることができる:

$$\min_B \sum_{\ell} \frac{1}{\mathcal{M}} \sum \left(P(X, Y) - \|B(X - Y)\|^2 \right). \quad (3)$$

自然言語分野では、類似度を、ユークリッド距離ではなく、コサイン類似度 $((x \cdot y) / (|x||y|))$ で定められることが仮定される場合がある。単語 (あるいは項目) x が与えられた場合に単語 y の想起確率を次式で定義する:

$$P(y|x) = \frac{e^{\text{Sim}(y,x)/\beta}}{\sum_{z \in \mathcal{Z}_{t-\tau:t}} e^{\text{Sim}(z,x)/\beta}} \quad (4)$$

ここで β は温度パラメータである。 $\beta \rightarrow 0$ であれば決定論的に振る舞うことになる。上式 (4) 分母の総和範囲 $[t - \tau : t]$ における τ は短期記憶の範囲を表すものと解される。健常者であれば、自由想起系列中に一旦表出した語は表出することがない。だが、健忘症例では場合によって考慮が必要な場合があると思われる。

このとき、人間の行う想起確率と P_h と単語埋め込みモデルから算出される想起確率 P_m との間の KL ダイバージェンスを目的関数とする学習が定義できる [5]。

$$D_{\text{KL}}(P_h \| P_m) = \sum_{i,j \in \mathcal{D}} P_h \log \left(\frac{P_h}{P_m} \right) \quad (5)$$

上述の最小自乗基準、あるいは、KL 基準を用いて、単語埋め込みモデルと人間の課題遂行データを比較することが可能となる。

2. 実験

2.1 材料

■意味表現 日本語ウィキペディア全文*1を MeCab*2 を用いて分かち書きにした。得られた全単語について、word2vec[7] で訓練した結果を用いた。訓練に用いたモデルは CBOW[6] でウィンドウサイズを 20 とし、埋め込み次元数を 200 とした。従って得られた単語の意味表現は各語とも 200 次元のベクトルとなる。

*1 <http://dumps.wikimedia.org/enwiki/latest/enwiki-latest-pages-articles.xml.bz2>

*2 <https://taku910.github.io/mecab/>

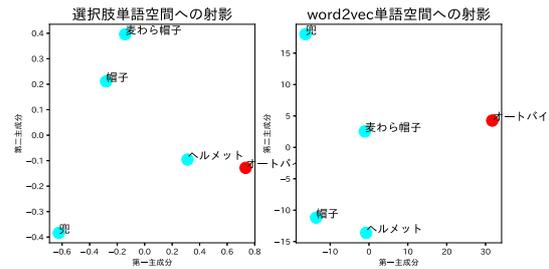


図 5: 4 肢選択課題の例。左: 選択肢で構成された部分空間への射影。右: 元の埋め込みベクトルの布置。課題: 兜, 麦わら帽子, 帽子, ヘルメット, の中から, ターゲット語: オートバイ, にもっとも近いと思われるものを選べ

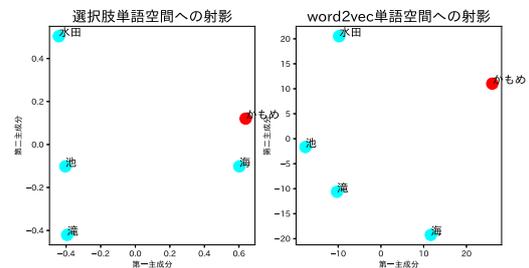


図 6: 課題: 水田, 池, 滝, 海, の中から, ターゲット語: かもめ, にもっとも近いと思われるものを選べ

2.2 4 肢選択課題への応用

手がかりを与えられ、4つの選択肢の中から最も近いものを選ぶようなピラミッドパームツリー検査 (PPTT) を考える。このとき、与えられた4つの選択肢からなる意味空間へ射影行列を定義し、この部分空間への射影したベクトル間の布置 (図 5 左) と、もとのベクトル間 (同図右) との比較を示した。

図 5 では、‘オートバイ’に対して選択肢に挙げられる4つの頭を覆う帽子類について、図右のように、通常の word2vec 空間では傾向が認められない。ところが、4つの帽子類で構成される空間へ‘オートバイ’を射影することにより、‘ヘルメット’との関係が明確となっている。

図 6 では、‘かもめ’に対する、水に関する景観‘水田’、‘池’、‘滝’、‘海’で構成される空間へと射影することにより、‘海’と‘かもめ’との関係が word2vec による空間附置よりも、射影空間の方が明確となる。図 7 では、正解であるターゲット語‘急須’と‘湯呑’との関係は word2vec 空間でも同様に近いが、射影空間ではコーヒーカップの附置が明確に分離されていることが見て取れる。図 8 でも、上 3 例と同様で、4つの建物選択肢空間へ射影した場合、‘黒板’と‘学校’とが明確に近づいている。

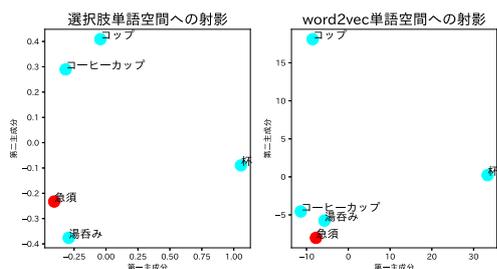


図7: 課題: コップ, コーヒーカップ, 湯呑み, 杯, の中から, ターゲット語: 急須, にもっとも近いと思われるものを選べ

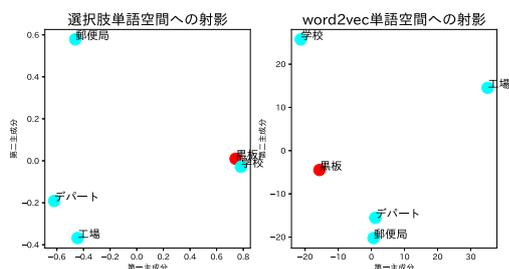


図8: 課題: 学校, 工場, デパート, 郵便局, の中から, ターゲット語: 黒板, にもっとも近いと思われるものを選べ

2.3 単語連想課題 1

たとえば, 神経心理学検査の一つに, 動物や野菜の自由想起, および「か」で始まる単語を思いつく限り表出する課題がある。この語連想課題を, 射影行列を用いた提案手法により模倣することを試みた。式(4)式に従って, 自由連想を行った例を以下に示す:

手がかり語: イヌ → ['カワウソ', 'ワニ', 'ザリガニ', 'ザリガニ', 'ウミヘビ', 'カワウソ', 'ニホンアナグマ']

オリジナル word2vec による イヌ の最近隣語 → ['イヌ', 'ネコ', 'オオカミ', 'ブタ', 'ウサギ', 'ウシ', '犬', 'アライグマ', 'ニワトリ', 'ハツカネズミ']

手がかり語: サンショウウオ → ['ザリガニ', 'アリクイ', 'ワニ', 'ヘビ', '猛禽類', 'アカギツネ', 'ザリガニ']

オリジナル word2vec による サンショウウオ の最近隣語 → ['サンショウウオ', 'タイコウチ', 'カナヘビ', 'アオガエル', 'は虫類', 'イモリ', 'アカハライモリ', 'オオサンショウウオ', 'アマガエル', 'ゲンゴロウ']

手がかり語: アニメ → ['ネコ', 'ザリガニ', 'ワニ', 'ヘビ', '猛禽類', 'アカギツネ', 'ザリガニ']

オリジナル word2vec による アニメ の最近隣語 → ['アニメ', 'テレビアニメ', 'アニメ版', '実写',

'OVA', '本編', '漫画', 'キャラクター', 'アニメーション', 'アニメ化']

上記想起語は, 式(4)に従って, 項目間類似度を確率密度に変換し, サンプルした結果である。与式中のパラメータ β を操作することにより, 決定論的な単語生成系列を得るか, もしくは, 多様な単語生成系列となるかを制御可能である。上記, 自由連想課題における単語産出例では, 既出の単語を抑制する処理を行っていない。この影響から, 同一単語が繰り返しサンプリングされることが起こっている。「ザリガニ」と連続して表出している。健常者であれば, このような想起語の連続は起こり得ないと予想される。種々の記憶障害, 痴呆症, あるいは保続と言った神経心理学的症状を想定した場合には, 既述の産出例のような想起系列が考えられると思われる。

2.4 単語連想課題 2

水野らは健常者に対して日本語 300 語を用いて対応する連想語を調べた [13]。彼女らのデータを説明することを試みた。水野らの用いた刺激語 300 語のうち 2.1 節で用いた word2vec に存在する 296 語の単語を用いた。'あたま', 'しにせ', 'だんご', 'はんこ', の 4 語は word2vec に登録が無いので除外した。これらの 4 単語は, 水野らが, ひらがな表記語刺激として用いた単語である。水野らのデータには, ターゲット語に対して想起された単語の総頻度で除した連想強度が記されている。水野らの連想強度と word2vec の最近接語との関係を次表に示す:

表 1

条件	単語数	割合
top1 が同じ	10	0.034
top10 以内に水野 top1 が存在	17	0.057
top100 以内に水野 top1 が存在	50	0.169
100 以外に水野水野 top1 が存在	142	0.480
上記以外の順位	77	0.260

2.5 オノマトペ

[15] から, 索引項目 2761 語のうち, 日本語ウィキペディアによる word2vec に登録項目のある 1761 語を用いて, オノマトペ空間を作成した。

■基本語彙 日本語を読むための語彙データベース *3 から, 使用範囲狭小語彙データベース *4 に記載されている 18406 語を用いた。

基本単語からオノマトペ空間での隣接語の検索例を

*3 <http://www17408ui.sakura.ne.jp/tatsum/database.html>

*4 http://www17408ui.sakura.ne.jp/tatsum/database/VDRJ_Ver1_1_Research

以下に示す (カッコ内はコサイン類似度):

1. 球茎 : (びっしり:0.332) (ぬめり:0.317) (でこぼこ:0.281) (ギシギシ:0.276) (ざらざら:0.259) (ひらひら:0.246) (キチン:0.244) (ピン:0.232)
2. 吐露 : (うんざり:0.551) (やきもき:0.480) (がっかり:0.457) (ヤキモキ:0.445) (ウンザリ:0.425) (あっけらかん:0.423) (ギクシャク:0.413) (きっぱり:0.409)
3. サイエントロジー : (ザック:0.249) (トロン:0.235) (ペロン:0.212) (ドロン:0.197) (クラッ:0.186) (うんざり:0.183) (バシリ:0.174) (ごたごた:0.172)

反対に、オノマトペから基本単語の検索結果を以下に示す。

1. ネットネット : eye(0.286) エンタメ (0.279) momo(0.277) スターデジオ (0.273) honey(0.266) remission(0.254) 多謝 (0.253) 漢薬 (0.251)
2. ケロケロ : パピ (0.478) ニャロメ (0.461) ポーポ (0.457) ロベ (0.453) イガグリ (0.442) アラモード (0.441) ロリポップ (0.439) バーバビカリ (0.438)
3. さっぱり : タレ目 (0.461) 朴念仁 (0.412) 下手 (0.399) 染め直し (0.390) ナヨナヨ (0.385) カッコイイ (0.369) 日持ち (0.368) 飲み手 (0.365)

2.6 オノマトペ生成

上記、オノマトペ 1761 をリカレントニューラルネットワーク [2, 4] を用いて単語表現を学習した。学習に用いたパラメータを用いてオノマトペ生成器を作成した。以下に算出例を示す。

ビククラ, パンパリ, チョッ, チャヘキ, オイオイ, サクサク, フクフコ, ガクラク, フクフコクン, コンタン, コッコリ, ガッキン, キッキリ, キッ, キッキリ, ウヤヘヤ, オンオン, キンムン, ポーン, ポコボソ, ポンベリ, マヨ, ヨヤガク, ヨチャクシャクシャ, キックリ, ギリッ, キワッ, ギリガグ, イクイク, クラッ, クタクン, ガッチリ, ガッコン, キッキリ, キワムワ, ムキムム, ポヤポタ, ポヨポヨ, ポリポリ, ポロポロ, ポロリ, ムン, コンボン, キー, キロウロ, ソロリ, ギリギリ, ゲッソリ, モッキン, キー, コロコロ, ワチョロ, ウロウネ, ウゴウハ, ウロウネチョロ, ワンサン, キンキラ, ポキン, ワッ, ムッ, ムンポリ, キワッ

3. 考察

適切な射影空間を定義することで、4 肢選択課題、単語想起課題、オノマトペ生成課題で示したような結果を得た。今後は、心理空間への変換を式 (3) あるいは式 (??) を用いることで変換することで、心理モ

デルとしての妥当性を検証する必要がある。その他にも、(1) ELMo[10] や BERT[1] などより性能の高い単語埋め込みモデルを用いる。(2) 訓練データセットの拡大を図る。今回用いた日本語 wikipedia 全文のデータと英語版 wikipedia のデータサイズを比較すれば、およそ 10 倍の違いがある。(3) 射影空間を適切に設定する。オノマトペの意味情報から、カタカナ語の部分情報を部分射影空間として削除し、音韻情報との共有空間へと射影するなど、の改善策が考えられる。これらの方向性は、自然な拡張のように思われる。

文献

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*, 2018.
- [2] Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.
- [3] John Hewitt and Christopher D. Manning. A structural probe for finding syntax in word representations. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4129–4138, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [4] Michael Irving Jordan. Serial order: A parallel distributed processing approach. Technical report, University of California, San Diego, San Diego, CA, May 1986.
- [5] Solomon Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951.
- [6] Tomas Mikolov, Quoc V. Le, and Ilya Sutskever. Exploiting similarities among languages for machine translation. *arXiv:1369.4168*, September 2013.
- [7] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Zoubin Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc., 2013.

- [8]Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies NAACL*, Atlanta, WA, USA, June 2013.
- [9]Jeffrey Pennington, Richard Socher, and Christopher D. Manning. GloVe: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, Oct. 2014.
- [10]Matthew E. Peters, Mark Neumann, Mohit Iyyery, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *arXiv preprint*, 2018.
- [11]Miloš Radovanović, Alexandros Nanopoulos, Marienburger Platz, and Mirjana Ivanović. Hubs in space: Popular nearest neighbors in high-dimensional data. *Journal of Machine Learning Research*, 11:2487–2531, 2010.
- [12]Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*,. OpenReview.net, 2019.
- [13]水野 りか, 清河 幸子, 川上 正浩, and 柳谷 啓子. **連想語頻度表 3 モーラの漢字・ひらがな・カタカナ表記語**. ナカニシヤ出版, 東京, 2011.
- [14]坂本 真樹. **感性情報学 -オノマトペから人工知能まで-**. コロナ社, 東京, 2018.
- [15]小野 正弘, editor. **日本語オノマトペ辞典**. 小学館, 東京, 第一版 edition, 2007.