

日本語版拡散的連想課題の改良

Refining the Japanese version of divergent association task

石黒 千晶¹、鈴木 崇太朗²、服部 雅史³、安倍 梨沙³、楊 鯤昊⁴、内海 彰⁵
Chiaki Ishiguro¹, Sotaro Suzuki², Masasi Hattori³, Lisa Abe³, Kunhao Yang⁴, Akira Utsumi⁵

¹東京大学、²金沢工業大学、³立命館大学、⁴芝浦工業大学、⁵電気通信大学
The University of Tokyo¹, Kanazawa Institute of Technology², Ritsumeikan University³,
Shibaura Institute of Technology⁴, The University of Electro-Communications⁵
ishigurochiaki37@gmail.com

概要

拡散的思考 (DT) の自動スコアリング手法として注目される意味的距離を用い、日本版拡散連想課題 (DAT-J) を開発し、その妥当性を検証した。日本語コーパス由来の単語ベクトルでスコアを算出し、3 回の調査により信頼性・妥当性を確認した。回答検証機能付き Web アプリも開発し、回答形式の違いがスコアに影響しないことを示した。さらに、大規模言語モデルの埋め込み層を用いたスコアリングも試みたが、精度の大幅な向上には至らなかった。

キーワード：創造性(creativity)、拡散的思考(divergent thinking)、自動採点(automated scoring)、意味的距離(semantic distance)

1. はじめに

拡散的思考 (divergent thinking: DT) は、課題に対して多様な解決法を生み出す力と定義され[1]、代表的な DT 課題である Alternative Uses Task (AUT) を用いて測定されてきた。AUT は、新聞やレngaなどのものの通常とは異なる使用法を考える課題であり、この課題への回答は採点者によって流暢性・柔軟性・独自性・精緻性などの観点で評価される。近年、この採点にかかる膨大なコストを削減するために、自然言語処理を用いた自動採点が提案されている[2]。

潜在意味分析 (LSA) では、言語コーパスから単語間の類似度を計算することできる。AUT の自動採点ではこの計算を応用して、アイデア同士やオブジェクトとアイデアの意味的距離を算出することで独自性や柔軟性を採点することができる[2]。Beaty & Johnson (2021) は、AUT の英語回答を自動採点するオンラインプラットフォームも開発し、他の言語にもこうした自動採点の試みは広がりつつある[3]。Olson et al. (2021) は意味的距離を用いたより簡便な DT 課題として拡散連想課題 (divergent association task: DAT) を提案した[4]。この課題は、お互いに意味や使い方が異なる名詞を 10 個回答

するというもので、数分程度の短時間で実施可能で、かつ、即時に自動採点することが可能で大規模データでもその妥当性や信頼性が示されている。DAT は英語以外の言語にも広がり[4, 5]、日本語版開発も始まっている。石黒・鈴木 (2022) は、英語版の例を参照して日本語版 DAT を作成し、同様のスコアが算出できることを示した[7]。さらに、石黒・鈴木・服部・安倍・楊(2025) は研究利用可能な 12 種類の単語ベクトル(word vectors: WV) を用いて、既存の DT 課題 (SA 創造性検査) や創造性関連指標 (性格特性や創造活動など) との関連を検討し、日本語版 DAT の基準関連妥当性を検討した[8]。石黒他(2025)の結果では、日本語版 DAT (以降は DAT-J) は WV によっては既存の DT 課題や創造性関連指標と有意な正の相関関係が見られ、日本語での DAT-J の使用可能性が確認された。ただし、DAT-J と DT 課題との相関関係は弱い相関にとどまっており、原版の DAT ほどの妥当性を持っているとは言い難い[8]。

そこで本研究では、日本語版 DAT (以降は DAT-J) の改良を目的として、日本語 WV 以外に、大規模言語モデルを用いた DAT スコアの算出を試みる。近年は WV だけでなく大規模言語モデルを用いた DT 課題の自動採点が行われている。DAT は単語同士の意味的距離を用いた採点を行うが、大規模言語モデルの embedding 層から取り出した WV を利用して単語スコアの計算も可能である。そのため、本研究では大規模言語モデルを用いた DAT-J スコアの妥当性について検討する。

2. 方法

本研究は石黒他(2025)で取得した以下のデータをもとに大規模言語モデルを用いた DAT スコア算出を行う。データに関する詳細は石黒他(2025)を参照されたい[8]。
参加者 芸術を専門としない大学生 358 名を対象に、

¹ 本研究は石黒・鈴木(2022)による探索的研究を発展させ、DAT-J の妥当性検証を行なった石黒・鈴木・服部・安倍・楊(2025)のデータを用いて、大規模言語モデルを用いた DAT-J の改良を試みたものである。

既存の DT 課題である SA 創造性検査と DAT-J を実施する集団実験と質問紙調査を実施した。本調査は第 1 著者の所属機関の倫理審査委員会の承認を受けた上で実施した。参加者は研究方法や手続きについて口頭、紙面で説明を受けた上で、研究参加に同意した場合のみ実験・調査に参加した。

調査手続き 参加者は、本調査のために独自に作成した DAT-J のウェブアプリにアクセスした。画面には「言葉の意味や使い方ができるだけ異なる単語を 10 個考えて入力してください。その際、以下のルールを守ってください」という教示が記載されていた。ルールは(1)日本の単語のみを入力すること、(2)名詞のみを入力すること、(3)固有名詞は入力しないこと、(4)特殊な用語は使わないこと、(5)身の回りのものを見たりせずに、自分で単語を考えることであった。参加者はこの教示に基づき 10 個の単語を回答し、性別や年齢、回答時に考えていたことなどを自由記述で回答した。

また、既存の拡散的思考課題として SA 創造検査 A 版 [9] を実施した。SA 創造性検査には AUT が含まれるため、本研究では AUT の回答データのみを研究に用いた。収集された回答は検査会社が流暢性(適切な回答の数)、柔軟性(回答のカテゴリの数)、独自性(集団内で出現頻度が低い回答の数)、精緻性(回答の具体性)の観点で採点した。

さらに、創造性に関連する心理尺度についてウェブアンケートで回答を求めた。アンケートは複数の尺度から構成されていた。創造性に関する性格特性を測定するため、短縮版ビッグファイブパーソナリティ尺度 [10] から経験への開放性、また、日本語版創造的自己尺度 [9] の 11 項目を用いて創造的自己について、7 件法(「1: 全くそう思わない」 - 「7: 強くそう思う」)で回答を求めた。

さらに、創造的行動や活動・達成など、パフォーマンスに関わる自己報告も求めた。創造的行動は日本語版創造的行動尺度 [10]を用いて、過去 12 ヶ月間で「短編小説を書いた」「実用品を発案して、作った」など 34 の創造的行動を行なったかどうかを回答するように求めた。創造的活動・達成の測定には日本語版創造活動・達成尺度 [11] を用いた。この尺度は文学、音楽など 8 の創造領域それぞれについて、過去 10 年間で実施した創造活動の回数、各領域で達成したレベルを回答してもらった。例えば、文学領域での創造活動は「短い文学作品(詩、短編小説など)を書いた」などの項目について「0: 一度もない」「1: 1-2 回」「2: 3-5 回」「3: 6-10

表 1. DAT-J スコアと SA 創造性検査の相関係数

	Flu	Flx	Org	Ela
Asahi.CB	.10	.19 **	.16 **	.07
Asahi.CB.r	.07	.16 **	.14 *	.03
Asahi.GV	.13 *	.19 **	.15 **	.15 **
Asahi.GV.r	.10	.17 **	.16 **	.14 *
Asahi.SG	.09	.19 **	.18 **	.09
Asahi.SG.r	.06	.18 **	.17 **	.06
chiVe	.02	.12 *	.11 *	.06
fastText	.07	.14 *	.16 **	.09
hottoSNS	.02	.15 **	.12 *	.07
Shiroyagi	.04	.12 *	.14 *	.00
Tohoku.A	.06	.14 *	.13 *	.05
Tohoku.E	.06	.18 **	.19 **	.10
OpenAI_small	.09 +	.12 *	.11 *	.13 *
OpenAI_large	.05	.09 +	.09 +	.07
OpenAI_ada	.11 *	.15 **	.09 +	.11 *
LLM-jp	.05	.09 +	.11 *	.02
LLM-jp (emb)	.05	.14 *	.14 *	.06
Ruri	.07	.14 *	.08	.07
Ruri (emb)	.02	.05	.02	.02

註) 相関係数はスクリーニング後のデータをリストワイズ除去して算出 (n = 257), Flu: Fluency, Flx: Flexibility, Org: Originality, Ela: Elaboration, emb は大規模言語モデルの embedding 層を用いたことを示す, + $p < .10$, * $p < .05$, ** $p < .01$

回」「4: 10 回より多い」から一つを選択するように求めた。創造的達成については「0: 私はこの分野に関わったことがない」「1: この分野の創作活動を一度やってみたことがある」「2: この分野で少なくとも 1 つはオリジナル作品を作成したことがある」などから「10: この分野で自分の作品を販売したことがある」など 11 の段階で当てはまるものをすべて回答してもらった。

3. 結果

Olson et al. (2021)に従い、DAT 回答時の自由記述から DAT の教示を誤解していた回答者はスクリーニングした上で DAT スコアと SA 創造性検査および創造性

表 2 DAT スコアと創造性関連尺度の相関係数

	CS	E	A	C	N	O	CB	CAct	CAch
Asahi.CB	.08	.02	.02	-.13	-.11	.10	.07	.04	.16 *
Asahi.GV	.08	.06	.02	-.04	-.08	.15 *	.09	.00	.06
Asahi.GV.r	.17 **	.03	.09	-.02	-.08	.17 *	.08	.13 *	.12 *
Asahi.SG	.15 **	.03	.02	-.10	-.09	.14 *	.10	.15 *	.10
chiVe	.12	.04	.00	-.08	-.10	.16 *	.05	.10	.17 *
fastText	.20 **	.09	.04	-.12	-.12	.23 ***	.10	.14 *	.16 *
hottoSNS	.19 **	.13 *	.02	-.19	-.06	.19 *	.16 *	.18 *	.23 *
Shiroyagi	.17 *	.10	.04	-.09	-.06	.15 *	.08	.15 *	.17 *
Tohoku.A	.17 **	.09	.02	-.09	-.06	.15 *	.07	.19 **	.19 *
Tohoku.E	.16 **	.14	.10	-.06	-.09	.15 *	.07	.13 *	.10
OpenAI_small	.06	.00	.07	-.01	-.11	.08	-.03	-.02	.09 +
OpenAI_large	.11	-.06	.12 *	-.07	-.07	.17 **	.04	.03	.22 ***
OpenAI_ada	.12 *	.02	.12 *	-.06	-.13	.17 **	.03	.01	.11 *
LLM-jp	.18 **	-.04	.00	-.13	-.15	.25 ***	.12 *	.14 *	.25 ***
LLM-jp (emb)	.12 *	-.03	.10 +	-.06	-.13	.13 *	.09 +	.11 *	.20 **
Ruri	.09 +	-.01	.03	-.05	-.10	.15 *	.06	.08 +	.13 *
Ruri (emb)	.06	.01	.01	.00	.16	.05	.03	-.07	.03

註) CS: Creative self, E: Extraversion, A: Agreeableness, C: Consciousness, N: Neuroticism, O: Openness, CB: Creative Behavior, CAct: Creative activities, CAch: Creative Achievements. 相関係数はスクリーニング後のデータをリストワイズ除去して算出(n = 251), emb は大規模言語モデルの embedding 層を用いたことを示す, + $p < .10$, * $p < .05$, ** $p < .01$

関連尺度の相関を算出した。相関係数は Olson et al. (2021)に従って片側検定を用いた [4]。なお、本研究では石黒他(2025)の WV を使った DAT スコアとの比較をするため、日本語 WV の結果と大規模言語モデルを用いた結果を併記した。DAT スコアを算出するために用いた日本語 WV、および、大規模言語モデルやその埋め込みモデルは付録に示した。

DAT スコアと SA 創造性検査 DAT スコアと SA 創造性検査の AUT の評価観点のスコアの相関係数を算出した (表 1)。日本語 WV を用いた DAT スコアと AUT の柔軟性と独自性の間には概ね有意な正の相関が見られた。大規模言語モデルを用いた場合にも OpenAI_small, ada, LLM-jp (emb), Ruri において柔軟性との正の相関が見られ、OpenAI_small, LLM-jp, LLM-jp (emb)の DAT スコアと独自性に正の相関が見られた。

DAT スコアと創造性関連尺度 DAT スコアと創造性関連尺度の相関係数を算出したところ (表 2)、創造的自

己、経験への開放性、創造的行動、創造的活動・達成において有意な正の弱い相関関係が見られた。特に、Asahi.GV.r, fastText, hottoSNS, LLM-jp では一貫してこれらの創造性に関わる特性や経験との正の相関関係が見られた。

4. 考察

本研究は、DAT-J の改良を目的として、日本語の既存 WV、および、大規模言語モデルを用いた DAT スコアの算出を行なった。その結果、既存の AUT の柔軟性や独自性スコアと大規模言語モデルを用いた DAT スコアは OpenAI_small, ada, LLM-jp (emb), Ruri などモデルによっては正の相関関係があることがわかった。また、創造性に関わる特性や経験を測る創造性関連尺度においても LLM-jp では一貫した正の相関関係が見られた。これらの結果を踏まえると、本研究で検討した範囲では LLM-jp を用いて算出した DAT スコアは、既存の DT 課

題や創造性関連指標の両方でバランスよく妥当性が見られたと言える。ただし、DAT スコアと既存の DT 課題との相関はいずれも弱い効果量にとどまり、DAT-J が測定する拡散的思考は既存の DT 課題と必ずしも一致しないかもしれない。原版 DAT と比較すると、DAT-J の基準関連妥当性には依然として課題が残る。

今後は、DAT-J のさらなる改良を目指して、より DAT-J に適したモデルや学習データ、日本語に抵抗した大規模言語モデルの選定を検討する必要がある。今後、より多様なモデルやファインチューニングされた日本語特化モデルを用いて、単語の意味的距離にフォーカスした DAT-J への適用方法を検討する必要がある。

文献

- [1] Guilford, J. P. (1967). The nature of human intelligence. McGraw-Hill.
- [2] Acar, S., & Runco, M. A. (2019). Divergent thinking: New methods, recent research, and extended theory. *Psychology of Aesthetics, Creativity, and the Arts*, 13(2), 153–158. <https://doi.org/10.1037/aca0000231>
- [3] Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods*, 53(2), 757–780. <https://doi.org/10.3758/s13428-020-01453-w>
- [4] Olson, J. A., Chmoulevitch, N. D., Cropper, S. J., & Webb, M. E. (2021). Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences*, 118(25), Article e2022340118. <https://doi.org/10.1073/pnas.2022340118>
- [5] Puvia, E., Taibi, D., & Tressoldi, P. (2022). The role of masked solutions in the accuracy of an insight problem-solving task. *Journal of Anomalous Experience and Cognition*, 2(1), 144–157. <https://doi.org/10.31156/jaex.23304>
- [6] Hubert, K. F., Awa, K. N., & Zabelina, D. L. (2024). The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Scientific Reports*, 14(1), Article 3440. <https://doi.org/10.1038/s41598-024-53303-w>
- [7] 石黒 千晶・鈴木 崇太朗 (2022). 拡散的思考課題の採点自動化の試み: Divergent Association Task 日本語版の開発 日本電子情報通信学会技術報告, 122(166), HCS2022- 39, 18–20.
- [8] 石黒千晶, 鈴木崇太朗, 服部雅史, 安陪梨沙, & 楊鯤昊. (2025). 日本語版拡散連想課題の開発と妥当性. *認知科学*, 32(1), 41-55.
- [9] 創造性心理研究会 (編) (1969). S-A 創造性検査 (A 版) 東京心理
- [10] 小塩 真司・阿部 晋吾 (2012). 日本語版 Ten Item Personality Inventory (TIPI-J) 作成の試み パーソナリティ研究, 21 (1), 40–52. <https://doi.org/10.2132/personality.21.40>
- [11] Ishiguro, C., Matsumoto, K., Agata, T., & Okada, T. (2022). Development of the Japanese Version of the Short Scale of Creative Self. *Japanese Psychological Research*. <https://doi.org/10.1111/jpr.12418>
- [12] Ishiguro, C., Matsumoto, K., Agata, T., Noguchi, H., & Okada, T. (under review). Examining the Japanese Version of the Biographical Inventory of Creative Behavior among Adults and Children Using Item Response Theory.
- [13] Ishiguro, C., Matsumoto, K., Agata, T., & Okada, T. (2024). Development of the Japanese Version of the Inventory of Creative Activities and Achievements. *Thinking Skills and Creativity*, <https://doi.org/10.1016/j.tsc.2024.101648>

付録

- Asahi.CB/CB.r/GV/GV.r/SG/SG.r** 朝日新聞 WV(朝日新聞社 (株)レトリバ; 田口他, 2017). <https://cl.asahi.com/apidata/wordembedding.html>
- chiVe** chiVe: Sudachi と NWJC による日本語 WV(国立 国語研究所/(株)ワークスアプリケーションズ; 真鍋他, 2019). <https://github.com/WorksApplications/chiVe>
- fastText** Pre-trained word vectors for 157 languages (Facebook Inc.; Grabe et al., 2018). <https://fasttext.cc/>
- hottoSNS** hottoSNS-w2v: 日本語大規模 SNS+Web コーパスによる単語分散表現モデル((株)ホットリンク; 松野他, 2019). <https://github.com/hottolink/hottoSNS-w2v>
- Shiroyagi** word2vec 学習済み日本語モデル((株)白やぎ コーポレーション) <https://aial.shiroyagi.co.jp/2017/02/japanese-word2vec-model-builder/>
- Tohoku.A/E** 日本語 Wikipeda エンティティベクトル (Suzuki et al., 2018). <http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki/vector/>
- Open AI small/large/ada** OpenAI による多言語対応テキスト埋め込みモデル.それぞれ text-embedding-3-small / text-embedding-3-large / text-embedding-ada-002 に対応 <https://platform.openai.com/docs/guides/embeddings>
- LLM-jp/LLM-jp (emb)** 国立情報学研究所大規模言語モデル研究開発センターによる日本語特化 LLM・埋め込みモデル (llm-jp プロジェクト)llm-jp/llm-jp-3-1.8b. <https://llm-jp.notion.site/llm-jp-3-1-8b-ada277c2fc8a4df1a64050a23f3a6a99>
- Ruri / Ruri(emb)** 名古屋大学が開発した日本語大規模テキストコーパスに基づく単語埋め込みモデル(塚越・笹野, 2025). <https://ruri.llm.jp/>