単語の意味の通時変化とドメイン変化は区別できるのか

Tan Xin　木山朔　凌志棟　小町守
一橋大学
{xin, hajime, shito, komachi}@scl.sds.hit-u.ac.jp

概要

通時的な単語の意味変化とは，単語の意味や関連性が時間によって変化する言語学的事象である．意味変化の分析において単語埋め込みを用いた分析手法があるが，これは共起単語の変化を分析していることに相当する．共起単語の変化は，時間経過による言語学的な変化と，ドメインが変わることによる変化の 2 種類の要因が考えられる．従来の単語埋め込みによる分析ではこれら 2 つの変化をまとめて分析しているため，通時変化を分析しているのかドメイン変化を分析しているのかが明らかではない．本研究では，この「通時変化とドメイン変化は区別できるのか」という問いに取り組み，現行の意味変化の分析手法では「通時変化とドメイン変化が区別できない」ことを示す．

1　はじめに

単語の通時的な意味変化とは，単語の意味や関連性が時間経過によって変化する言語学的事象のことを指す [1, 2, 3]．例えば，record という単語は，主に「物事を記録する」という意味で用いられているが，現在では従来の意味に加えて「音楽を再生するメディア」を指す意味でも用いられている．このような単語の意味変化を計算的に分析する研究が，自然言語処理の分野では広く行われている．

意味変化の分析手法として単語埋め込みを用いた分析手法がある [1, 2, 4]．単語埋め込みは分布仮説 [5, 6]，つまり共起単語の変化を意味の変化とみなして分析を行うものである．ここで，共起単語の変化の主要な要因として，時間経過による言語学的な意味の変化と，ドメインの違いによって生じる文脈の変化が挙げられる．ドメインの定義は複数存在するが，本研究では，「文体的や機能的なカテゴリであるジャンル」として定義する [7, 8]．単語埋め込みの変化を分析することは，上記の主要な 2 種類の変化が混在した状態をまとめて分析していることとなり，その結果，通時的な変化を分析しているのか，ドメインの変化を分析しているのかを区別できないことが予期される．

本研究では，この「通時変化とドメイン変化は区別できるのか」という問題に対して，コーパス構成の観点から体系的に検証を行う．具体的には，英語の現代語コーパスである Corpus of Contemporary American English（COCA）を対象に，複数時期かつ新聞, 雑誌, 学術など 6 つのドメインに分類されたデータを用い，全体コーパスと各ドメイン別のコーパスの双方について意味変化の分析を行う．

本研究の分析は 2 つから構成される．第一の分析では，ドメインと時間の両方を考慮した単語埋め込みのコサイン類似度行列を作成し，「単語のドメイン内とドメイン間における通時的な変化に差がある」を定性的に説明する．第二の分析では，動的な単語埋め込みのクラスタリングを行い，「全体コーパスで観測される語義遷移と，各ドメインで観測される語義遷移が一致するのか」を定性的・定量的に分析する．これらにより，同一単語であってもドメインごとに単語埋め込みや語義遷移が大きく異なること，すなわち「通時変化とドメイン変化が区別できない」ことを示す．

2　関連研究

複数時期の意味変化　複数時期にかけて意味変化を分析する研究を紹介する．隣接時期に着目する研究 [9, 10, 11, 12, 13, 14] や，任意の時期間に着目する研究 [15]，BERT やトピックモデルを用いて語義の割合を可視化する研究 [16, 17, 18, 19]，意味変化のパターンに着目する研究 [20, 21] が行われている．どのように意味変化が起きるのかを分析するには，複数時期に渡った意味変化の分析手法が必要であると考え，本研究では複数時期の意味変化の分析手法を用いたドメイン間に渡った分析を行う．

意味変化の種類の識別　先行研究において「変化の種類の識別」を試みたものが存在する．Hamilton

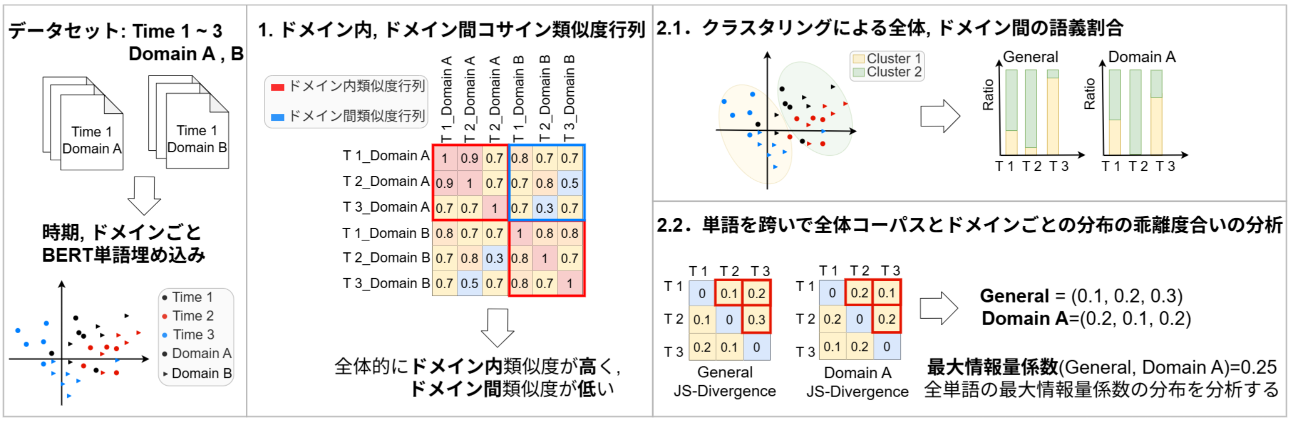

図 1　意味変化とドメイン変化の分析フレームワーク．各時期・ドメインの単語埋め込みを用いて (1) ドメイン内・間の類似度行列を分析し，(2) クラスタリングを行い全体と各ドメインの語義変化の相関を定量化する．

ら [22] は，単語埋め込みの変化を 2 つに分割し，言語内部の緩やかな変化と，文化的背景の変化を区別できる可能性を示した．しかし，この研究は「意味変化の内的性質」を分解するものであり，コーパス内のドメイン分布が時間とともに変動することによる外的要因としてのドメイン変化は明示的には扱っていない．つまり従来研究は，意味変化の計測値にドメイン変化がどの程度混入しているか，また両者が分離可能であるかという根本的な問題に対して十分な検証を行っていない．本研究では，同一コーパス内の複数のドメインと全体コーパスを比較することで，ドメイン変化を考慮した分析を行う．

3　分析フレームワーク

本研究では，各時期におけるドメイン間の語義の違いを比較し，意味変化の分析に混入しうるドメイン変化を可視化するためのフレームワークを提案する（図 1）．

3.1　類似度行列による分析

まず，対象単語集合を $\mathbb{W} = \{w_1, w_2, \ldots, w_{N_{\mathbb{W}}}\}$ とし，各用例に時期情報およびドメイン情報が付与されたコーパスを用意する．コーパスに含まれる時期の集合を $\mathbb{T} = \{t_1, t_2, \ldots, t_{N_{\mathbb{T}}}\}$，ドメインの集合を $\mathbb{D} = \{d_1, d_2, \ldots, d_{N_{\mathbb{D}}}\}$ とする．対象単語 w について，時期 t_i・ドメイン d_m における用例集合を $\mathbb{U}_{t_i}^{d_m} = \{u_{t_i,1}^{d_m}, u_{t_i,2}^{d_m}, \ldots, u_{t_i,N_{\mathbb{U}}}^{d_m}\}$ と定義する．ここで，$u_{t_i,k}^{d_m}(w)$ は時期 t_i，ドメイン d_m における k 番目の用例中の単語 w を表す．各用例 $u_{t_i,k}^{d_m}(w)$ に対し，文脈を考慮した単語埋め込み $e_{t_i,k}^{d_m}(w)$ を獲得する．

次に，対象単語ごとにコーパスから用例を抽出し，時期・ドメインごとに語義を表現する平均埋め込みを計算する（図 1-1）．時期 t_i・ドメイン d_m における平均ベクトルは

$$\bar{e}_{t_i}^{d_m}(w) = \frac{1}{N_{\mathbb{U}}} \sum_{k=1}^{N_{\mathbb{U}}} e_{t_i,k}^{d_m}(w)$$

として計算できる．続いて，通時的な意味変化の中に含まれるドメイン変化を捉えるため，各時期 t_i においてドメイン類似度行列を

$$\begin{aligned} &S_{\mathbb{T}}^{\mathbb{D}}(w)[N_{\mathbb{T}}(m-1)+i, N_{\mathbb{T}}(n-1)+j] \\ &= \mathrm{sim}(\bar{e}_{t_i}^{d_m}(w), \bar{e}_{t_j}^{d_n}(w)) \quad (m, n \in N_{\mathbb{D}}, i, j \in N_{\mathbb{T}}) \end{aligned}$$

により算出する．ここで $S_{\mathbb{T}}^{\mathbb{D}}(w)$ は，単語 w における時期集合 $\mathbb{T}$，ドメイン集合 $\mathbb{D}$ における通時的な単語埋め込みの平均のドメイン間の類似度を表す行列である．類似度 $\mathrm{sim}(\cdot,\cdot)$ にはコサイン類似度を用いる．行列の各行・各列は，「時期 $\times$ ドメイン」の組に対応しており，同一ドメイン内で異なる時期における類似度，ならびに異なるドメイン間での類似度を同時に表現している．この手法は木山ら [15] の類似度行列の分析手法にドメイン軸を加えたものとなる．

可視化された図（図 1-1）において，対角線上に配置されているブロックは，同一ドメイン内における通時的な類似度行列を表している．一方，対角線以外の上三角成分に現れるブロックは，ドメイン間の類似度行列に対応しており，異なるドメイン間で単語 w がどの程度類似した文脈で使用されているかを示している．したがって，対角線ブロックと非対角ブロックのコントラストから，ドメイン内の安定性とドメイン間の差異を直観的に把握できる．

3.2　語義分布による分析

コサイン類似度の計算に加えて，対象語の各ドメイン内および全体コーパス横断の意味変化を分析す

るために，教師なしクラスタリング手法 k-means を適用する（図 1-2）．ここでは，単語 w に対し，全時期，ドメインの単語埋め込み $e^{d_m}_{t_i,k}(w)$ を用いてクラスタリングを行う．

特定の単語に対する分析　ドメイン d_m における各時期 t_i のクラスタ割合を語義分布とみなし，その分布の推移から通時的な語義変化を分析する [16, 17]．特定の単語 w についてクラスタの解釈やドメインごとの語義分布の割合を可視化することで，ドメインごとの語義分布の違いについて分析する（図 1-2.1）．

全単語に対する分析　続いて，全単語 $\mathbb{W}$ に対して全体コーパスから計算した変化と各ドメインから計算した変化の近さを調べる．変化を表現するには，各時期間におけるクラスタの割合 の Jensen-Shannon ダイバージェンス（JSD）を計算する．得られた時期間の JSD は図 1-2.2 のような $N_{\mathbb{T}} \times N_{\mathbb{T}}$ の行列と表現できる．そして，全体コーパスと各ドメインの語義分布の近さを測るために，全体コーパスの JSD 行列と各ドメインの JSD 行列の対角以外の上三角成分の近さを最大情報量係数 (Maximal Information Coefficient; MIC) [23] で表現する．MIC は 2 つの変数間の線形および非線形な関連性の強さを測定するための統計的指標である．MIC によって，全体コーパスとドメイン d_m における語義分布の時間的変化が，線形・非線形を問わずどの程度強く対応しているかを評価できる．値が小さいほど，ドメイン固有の語義変化が全体傾向から乖離していることを示す．この手法を全対象単語に適用し，得られた各単語の MIC の分布を可視化することで，ドメインごとに対象単語の変化が全体コーパスとどれくらい乖離しているかを分析する．

4　通時変化とドメイン変化の分析

4.1　実験設定

データセットとモデル　COCA は 1990 年から 2019 年までの 30 年間にわたる英語テキストで構成され，Academic（学術），Magazine（雑誌），News（新聞），Fiction（フィクション），Spoken（話し言葉），TV/Movies（テレビ番組），Blog/Web [1] の 7 ドメインを含む．対象単語は，全ドメイン・全時期で出現頻度が 100 回以上の内容語 458 語を選定した．モデルは，文脈化された単語埋め込みモデル BERT [2] の最終出力層の 768 次元埋め込みを使用する．

1）ただし Blog/Web ドメインは個々のテキストの公開年が不明であるため分析から除外し，本研究では Blog/Web を除く 6 ドメインのデータを使用する．

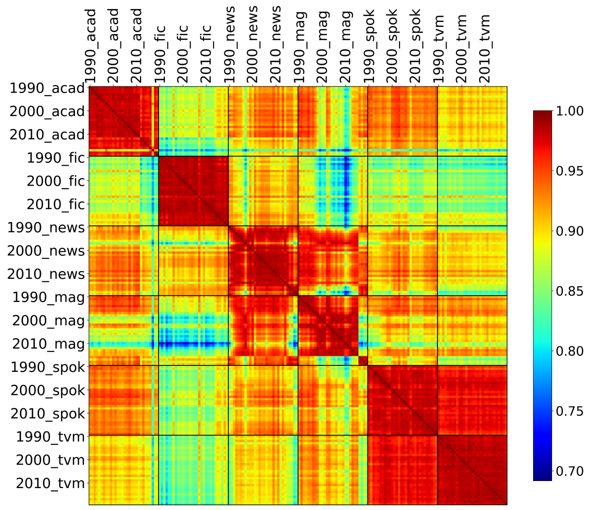

図 2　“Add” の時期 × ドメインのコサイン類似度行列

クラスタリング　単語の語義分布を推定するため，本研究では，単語埋め込みに対して k-means を適用する．各単語について得られた単語埋め込みをクラスタリングすることで，異なる語義に対応するクラスタを同定し，各クラスタの出現割合を語義分布として定義する．クラスタ数 k は $k = 1$ から 20 の範囲で k-means を実行し，シルエットスコアが最大となる値を最適なクラスタ数として採用する．

4.2　類似度行列の実験結果：Add

Add に関するドメイン内およびドメイン間の類似度行列（図 2）を分析した結果，雑誌と新聞ドメイン内では 2016 年以降に変化が見られる一方で，残りの 4 ドメインにおいて，30 年間にわたるドメイン内の類似度は全体的に高い水準を保っている．

ドメイン間の類似度に着目すると，フィクションと他のドメインとの間，特に雑誌ドメインとの間において，類似度の低下が最も顕著であることが確認される．これは，フィクションが有する文体的・語用的特性が，他のドメインとは大きく異なる文脈を形成しているためであると考えられる．また，テレビ番組ドメインと話し言葉ドメインの間では非常に高い類似度が観測されており，両ドメインが口語的で会話に近い言語使用を共有していることを反映していると解釈できる．以上の分析から，Add の単語埋め込みは，ドメイン内では概ね安定しているものの，ドメインや時期によって変化が生じており，ド

2）https://huggingface.co/google-bert/bert-base-uncased

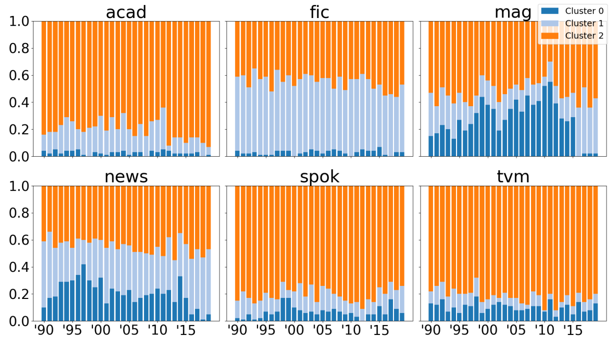

図 3 “Add” のドメインごと語義分布

メイン間では差異が存在することがわかる．

4.3 語義の分布の実験結果

定性分析：Add 対象単語 Add に対して，各ドメインにおける語義遷移を比較することで，ドメイン差が語義レベルでどのように現れるかを詳細に分析する（図 3）．1990 年から 2019 年にかけて，中心的な「何かを加えて量や規模を増やす」意味（クラスタ 2）は，フィクションおよび新聞を除くドメインにおいて最も高頻度で出現している．一方，「付け加えて述べる」意味（クラスタ 1）はフィクションと新聞で最も多く用いられ，「材料, 要素を加える」意味（クラスタ 0）は雑誌および新聞を除くドメインでは比較的まれである．

この結果は，学術文書では数量的意味が支配的であるのに対し，フィクションや新聞では会話的, 報道的慣習を反映した発話行為的意味が多用されること，また雑誌ドメインではレシピや DIY，ライフスタイル記事の多さを背景に，操作的意味の比率が相対的に高いことを示している．すなわち，各ドメインはジャンル固有の言語使用上のニーズに応じた特徴的な Add の用法を示していることが分かる．

また，ドメイン内における語義分布の変化には明確な差異が観察される．類似度行列の傾向と同様に，雑誌, 新聞ドメインでは顕著な意味変化が確認されている．雑誌ドメインにおいて 2016 年以降，操作的な意味，とりわけレシピ関連文脈における Add の使用はほぼ消失し，2019 年にはわずか 1 例のみが確認された．新聞ドメインでも 2016 年以降に同様の傾向は見られる．以上より，Add の語義変化の現れ方はドメインごとに異なり，ジャンル特有の使用慣習を反映した差異が存在している．

定量分析：単語を跨いだ傾向 本研究では，全体コーパスと各ドメインにおける意味変化の関係を，MIC を用いて定量的に評価した．MIC の値は，小さいほど両変数間の依存関係が弱く，大きいほど強い依存関係が存在することを示す．

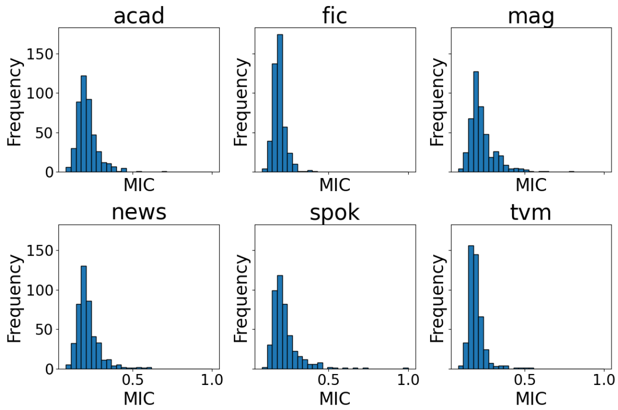

図 4 全体 vs. ドメインの最大情報量係数（MIC）

図 4 に示すように，全単語についてドメインごとに算出した MIC の値は 0.2 付近に集中しており，右歪みの分布を持っている．これは，多くの単語・ドメインの組み合わせにおいて，ドメイン固有の意味変化が，全体コーパスにおける意味変化の傾向とほとんど情報を共有していないことを示唆している．特に，フィクションおよびテレビ番組ドメインでは，MIC の値が 0.2 付近に集中しており，これらのドメインにおける意味変化は，全体コーパスとの関係性が比較的弱いと解釈できる．

一方で，一部のドメインでは比較的大きな MIC が観測され，全体的な意味変化傾向と強く連動した変化構造を持つことを確認できた．特に，雑誌ドメインにおける全単語の MIC の値の分布は右に裾が厚く，6 ドメインの中で最も全体コーパスとの意味変化傾向が関連していると考える．

5 おわりに

本研究では，単語の意味変化の分析において，通時変化とドメイン変化がどの程度区別可能であるのかを，コーパス構成の観点から検証した．COCA を用い，全体コーパスとドメイン別コーパスを比較した結果，単語埋め込みに基づく意味変化の指標は，時間に伴う語義変化のみならず，ドメイン分布の違いに強く影響されることが明らかになった．すなわち，従来の大規模コーパスを用いた意味変化の分析では，実際にはドメイン変化を反映している可能性が高く，現行の意味変化の分析手法ではこれら二つの変化が区別できないことが示された．今後の展望としては，通時変化とドメイン変化を分離してモデル化する手法の開発が挙げられる．

謝辞

本研究の一部は JST さきがけ JPMJPR2366 および JSPS 科研費 25H00470 の支援を受けたものである．

参考文献

[1] Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. Diachronic word embeddings and semantic shifts: a survey. In **Proceedings of the 27th International Conference on Computational Linguistics**, pp. 1384–1397, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.

[2] Francesco Periti and Stefano Montanelli. Lexical semantic change through large language models: a survey. **ACM Computing Surveys**, Vol. 56, No. 11, p. 1–38, June 2024.

[3] Jader Martins Camboim de Sá, Marcos Da Silveira, and Cédric Pruski. Survey in characterization of semantic change. **arXiv preprint arXiv:2402.19088**, 2024.

[4] Francesco Periti, Pierluigi Cassotti, Haim Dubossarsky, and Nina Tahmasebi. Analyzing semantic change through lexical replacements. In **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 4495–4510, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

[5] Zellig S. Harris. Distributional structure. **WORD 10 (2–3)**, 1954.

[6] John Rupert Firth. A synopsis of linguistic theory. **Studies in Linguistic Analysis (pp. 1-31). Special Volume of the Philological Society**, 1957.

[7] Marlies van der Wees, Arianna Bisazza, Wouter Weerkamp, and Christof Monz. What's in a domain? analyzing genre and topic differences in statistical machine translation. In Chengqing Zong and Michael Strube, editors, **Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)**, pp. 560–566, Beijing, China, July 2015. Association for Computational Linguistics.

[8] Alan Ramponi and Barbara Plank. Neural unsupervised domain adaptation in NLP—A survey. In Donia Scott, Nuria Bel, and Chengqing Zong, editors, **Proceedings of the 28th International Conference on Computational Linguistics**, pp. 6838–6855, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.

[9] Sunny Mitra, Ritwik Mitra, Martin Riedl, Chris Biemann, Animesh Mukherjee, and Pawan Goyal. That's sick dude!: Automatic identification of word sense change across different timescales. In **Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1020–1029, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[10] Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. Statistically significant detection of linguistic change. In **Proceedings of the 24th International Conference on World Wide Web**, p. 625–635, Republic and Canton of Geneva, CHE, 2015. International World Wide Web Conferences Steering Committee.

[11] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. In **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1489–1501, Berlin, Germany, August 2016.

[12] Feltgen Quentin, Fagard Benjamin, and Nadal Jean-Pierre. Frequency patterns of semantic change: corpus-based evidence of a near-critical dynamics in language change. **Royal Society Open Science**, 2017.

[13] Zijun Yao, Yifan Sun, Weicong Ding, Nikhil Rao, and Hui Xiong. Dynamic word embeddings for evolving semantic discovery. In **Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining**, p. 673–681, New York, NY, USA, 2018. Association for Computing Machinery.

[14] Philippa Shoemark, Farhana Ferdousi Liza, Dong Nguyen, Scott Hale, and Barbara McGillivray. Room to Glo: A systematic comparison of semantic change detection approaches with word embeddings. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 66–76, Hong Kong, China, November 2019. Association for Computational Linguistics.

[15] 木山朔, 相田太一, 小町守, 小木曽智信, 高村大也, 持橋大地. 通時的な類似度行列に基づく単語の意味変化の分析. 自然言語処理, Vol. 32, No. 4, pp. 1189–1240, 2025.

[16] Renfen Hu, Shen Li, and Shichen Liang. Diachronic sense modeling with deep contextualized word embeddings: An ecological view. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 3899–3908, Florence, Italy, July 2019. Association for Computational Linguistics.

[17] Mario Giulianelli, Marco Del Tredici, and Raquel Fernández. Analysing lexical semantic change with contextualised word representations. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 3960–3973, Online, July 2020.

[18] Seiichi Inoue, Mamoru Komachi, Toshinobu Ogiso, Hiroya Takamura, and Daichi Mochihashi. Infinite SCAN: An infinite model of diachronic semantic change. In **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**, pp. 1605–1616, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.

[19] 小林千真, 相田太一, 岡照晃, 小町守. BERT を用いた日本語の意味変化の分析. 自然言語処理, Vol. 30, No. 2, pp. 713–747, 2023.

[20] Pierluigi Cassotti, Stefano De Pascale, and Nina Tahmasebi. Using synchronic definitions and semantic relations to classify semantic change types. In **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 4539–4553, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

[21] Naomi Baes, Nick Haslam, and Ekaterina Vylomova. A multidimensional framework for evaluating lexical semantic change with social science applications. In **Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1390–1415, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

[22] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. Cultural shift or linguistic drift? comparing two computational measures of semantic change. In Jian Su, Kevin Duh, and Xavier Carreras, editors, **Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing**, pp. 2116–2121, Austin, Texas, November 2016. Association for Computational Linguistics.

[23] David N Reshef, Yakir A Reshef, Hilary K Finucane, Sharon R Grossman, Gilean McVean, Peter J Turnbaugh, Eric S Lander, Michael Mitzenmacher, and Pardis C Sabeti. Detecting novel associations in large data sets. **science**, Vol. 334, No. 6062, pp. 1518–1524, 2011.

A　MIC の算出過程

MIC を用いて全体コーパスとドメイン d_m における語義分布の時系列変化が，どの程度相関しているかを測定する．MIC を計算する際に，JSD 行列の対角以外の上三角成分を行方向に展開し，一列のベクトルにして入力する．全体コーパスの JSD ベクトルを $\mathbf{X}^{*w}$，ドメイン d_m の JSD ベクトルを $\mathbf{Y}_m^w$ とし，これら二つのベクトル間の MIC は

$$MIC(\mathbf{X}^{*w}, \mathbf{Y}_m^w)(w) = \max_{ab \leq B(n)} \left\{ \frac{I^*(\mathbf{X}^{*w}, \mathbf{Y}_m^w)}{\log\left(\min\{a, b\}\right)} \right\}$$

を算出する．サンプルサイズ n のデータセット $D = \{(x_i, y_{m,i})\}_{i=1}^{n}$ に対し，まず X 軸を a 個，Y 軸を b 個のビンで分割する格子 $G \in \mathscr{G}_{a,b}$ を考える．各格子 G 上で計算される相互情報量（Mutual Information）を $I(\mathbf{X}^{*w}; \mathbf{Y}_m^w \mid G)$ とする．ここで，分子の $I^*(\mathbf{X}^{*w}, \mathbf{Y}_m^w)$ は，与えられた分割数 (a, b) において格子の境界線を最適化することで得られる相互情報量の最大値である．分母の $\log(\min\{a, b\})$ は正規化項であり，スコアを $0 \leq MIC \leq 1$ の範囲に制約する．最終的な MIC 値は，サンプルサイズ n に依存する上限関数 $B(n)$ の制約下で，$\frac{I^*(\mathbf{X}^{*w}, \mathbf{Y}_m^w)}{\log\left(\min\{a,b\}\right)}$ を最大化して算出される．本研究では，先行研究に基づき，制約条件として $B(n) = n^{0.6}$ を採用した [23].

B　Degree の分析

類似度行列の定性分析　図 5 から学術とフィクションドメインの間（図の一列目）では，他のドメインと比較して類似度が低く，語義分布が相対的に乖離している．一方，フィクション, 新聞，雑誌，話し言葉，テレビ番組の各ドメイン間では，概して高い類似度が保たれている．

ただし，新聞ドメインでは 2016 年から 2019 年にかけて顕著な変化が観察される．この変化は，新聞ドメイン内の類似度行列において中心を横切る十字型の低類似度領域として現れ，さらに類似度の低下がドメイン内にとどまらず行列全体に広がり，中心を貫く大きな十字型の低類似度構造として確認できる．これは，当該時期の新聞ドメインにおける Degree の用法が，それ以前の新聞記事のみならず，他ドメインの用法とも異なっていたことを示す．

語義分布の定性分析　図 6 に示すように，1990 年から 2019 年にかけて，学位を表す意味（クラスタ 3）は，テレビ番組・雑誌・フィクションといったドメインにおいて全体分布の中で最も高い割合を占めており，日常的に用いられる用法であると考えられる．一方，抽象的な程度・度合いを表す意味（クラスタ 0）は，すべてのドメインで多くの用例が確認され，とりわけ新聞や学術といった比較的フォーマルな文体において顕著である．温度や角度などの物理量を表す意味（クラスタ 1）は，全体として最も使用頻度が低いものの，新聞，話し言葉，テレビ番組では相対的に多く見られる．とりわけ話し言葉およびテレビ番組ドメインにおいて, クラスタ 1 は，長期にわたって主要な意味カテゴリとして安定した位置を占めている．さらに，犯罪の等級を表す意味（クラスタ 2）は，学術ドメインで最も多く出現し，次いで話し言葉ドメインで多く用いられている．

ドメイン内の通時的変化に着目すると，新聞ドメインでは 2016 年以降，温度・角度などの物理量を表す意味（クラスタ 1）が顕著に増加しており，報道内容や話題の変化が反映されている可能性が示唆される．こうした傾向は, 類似度行列において観察された変化とも整合する．

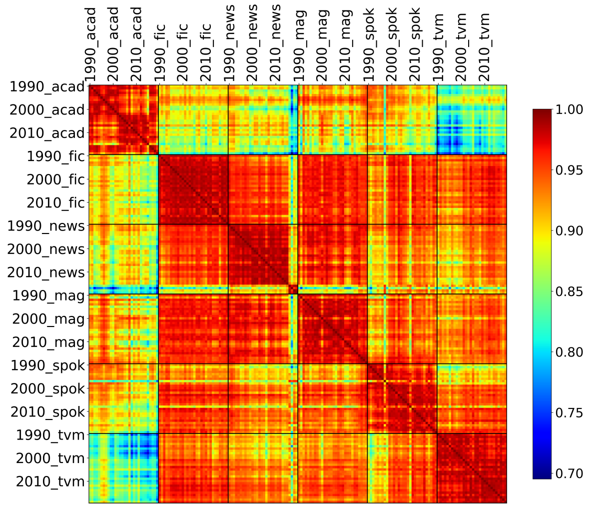

図 5　“Degree” の時期 × ドメインのコサイン類似度行列

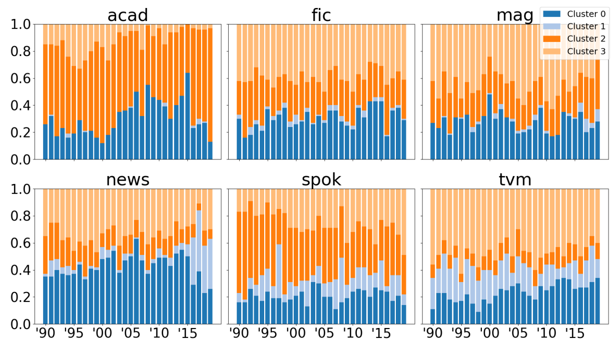

図 6　“Degree” のドメインごと語義分布