

出現頻度と意味変化に基づく建築学研究言説の分野間分析

向井 歩¹ 伊藤 真利子² 本間 裕大²

¹ 東京大学大学院 ² 東京大学 生産技術研究所

m-ayumi@g.ecc.u-tokyo.ac.jp

m-ito@iis.u-tokyo.ac.jp

yudai@iis.u-tokyo.ac.jp

概要

本研究は、2008年から2023年の日本建築学会大会梗概集 [1] を対象に、建築学研究における語の出現頻度と意味変化を定量的に分析し、分野間の言説構造の変遷を明らかにすることを目的とする。分野横断的に用いられる語の頻度推移をクラスタリングし、研究関心の長期的傾向と社会的事象に応答した変動を整理した。さらに、意味ベクトルを用いて語の意味的配置の年次変化を分析し、分野間の収束・乖離や意味変化の類型を抽出した。その結果、語の頻度変化と意味変化は必ずしも一致せず、社会的事象への応答が分野ごとに異なる形で意味構造に反映されることが示された。

1 はじめに

建築学研究のトピックは、災害・感染症・政策制度・技術進展といった外的刺激と相互に作用しながら変化してきた。2011年の東日本大震災は、耐震・防災技術の高度化に加え、復興計画やコミュニティ再建など生活世界の課題を計画学の主要テーマとして顕在化させた。また2020年以降の新型コロナウイルス感染症は、換気・空気質といった環境工学的課題と、学校・住宅・オフィス等の日常空間の運用・設計という計画・意匠的課題を同時に浮上させ、分野横断的議論を加速させた。さらに省エネ基準の強化や脱炭素政策の進展は、「炭素」「省エネ」「効率」「ZEB」などの単語を研究言説の中核へ押し上げ、環境性能が分野横断的な評価軸として定着する契機となった。

このような研究言説の変遷は、単語が「どれほど使われたか」(出現頻度)の変化だけでなく、同一単語が「どのような文脈で語られたか」(意味・用法)の変化としても現れる。したがって、分野横断的な知識再編を定量的に捉えるには、頻度推移と意味変

化を統合して扱う枠組みが必要である。近年、建築関連文献の電子化と自然言語処理手法の普及により、大規模テキストコーパスを対象とした通時的分析が現実的な選択肢となった。

本研究は、2008–2023年の日本建築学会大会梗概集を対象として、(i) 分野別の頻度推移、(ii) 頻度推移パターンの類型化、(iii) 意味ベクトル [2] に基づく分野間意味関係の可視化とその変化、を統合的に分析し、外的刺激が分野間相互作用を通じて知識体系を再編する過程を記述することを目的とする。

2 対象データと前処理

対象データは、日本建築学会大会梗概集 (2008–2023) に収録された「建築計画」「建築環境」「建築史・意匠」「建築構造」の4カテゴリである。PDFからページ単位で本文テキストを抽出し、ヘッダー・フッターおよび図表領域等レイアウト由来のノイズを除去したうえで、発表番号・分野内カテゴリ・タイトル等のメタ情報と紐づけてデータベース化した。

前処理として、表記正規化(全角半角・カタカナ等の統一)、ルールベース文分割、MeCabによる形態素解析(IPA辞書+建築・設備・都市計画分野の専門語を追加したユーザ辞書)、品詞選別(名詞・形容詞・副詞)、ストップワードおよび記号・混合文字トークンの除去を実施した。専門語(BIM, ZEB, CO₂, COVID-19等)を安定的に抽出しつつ、非言語記号やPDF由来ノイズを抑制することで、頻度推移分析と意味変化分析の両方に適したコーパスを構築した。

頻度推移分析では、低頻度語・分野固有語に起因する不安定性を抑え、分野横断的傾向に焦点を当てるため、分析対象単語を「上位共通語彙集合」として限定した。具体的に、(1) 4分野すべてで出現し、(2) 少なくとも1分野において16年累積出現回数が

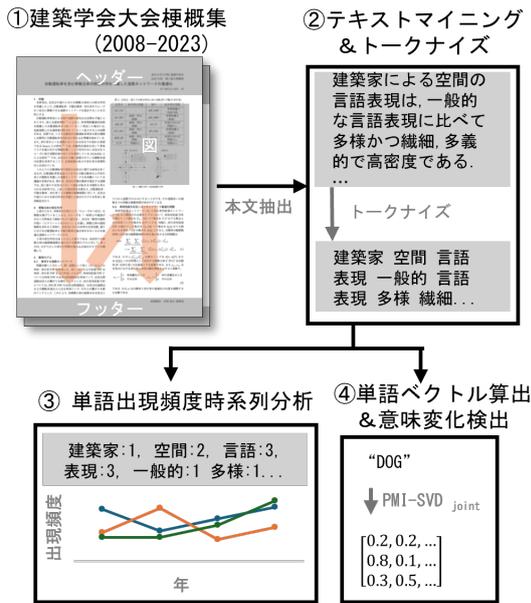


図1 本研究の分析プロセス

1000 回以上である単語を抽出し、3052 語を得た。
 上述したデータ取得に加えて、後述する分析手法を研究フローとして図 1 に示す。

3 手法

3.1 頻度推移の類型化

各分野・各年について単語の出現回数を集計し、「単語×(分野・年)」の行列を構築した。極端に高頻度な単語の影響を緩和するため $\log(1+x)$ 変換を行い、さらに単語ごとに L2 正規化して、絶対量の差ではなく「分野・年における相対的な分布形状」を比較可能にした。これを特徴量として k-means によるクラスタリングを行い、PCA 可視化と解釈可能性を踏まえてクラスタ数を 4 とした。頻度推移図(表 1)では、代表単語の時系列を示し、外的刺激(災害・感染症・政策変化)と対応するピークや増減の特徴を読み取る。

3.2 意味変化の類型化

意味変化分析は「計画、環境、歴史・意匠」の 3 分野を対象とする。これは、頻度分析を通じて建築構造分野が他分野と共有する単語割合が相対的に低く、分野間比較の安定性を確保しにくいことが確認されたためである。各年・各分野のコーパスから、対称ウィンドウ(±5 語)内の共起回数を計数し、全分野・全年を連結した共起行列を作成する。

これに PPMI 重み付けを施し、SVD により 300 次元の意味ベクトルを得る (PMI-SVDjoint[3, 4])。本手法は、学習後のアラインメントを必要とせず、全期間・全分野に共通する単一空間上で比較できる点に特徴がある。これにより、同一単語について「2008 年の計画」と「2021 年の環境」など、年・分野をまたぐ意味位置の比較が可能となる。

可視化として、意味ベクトルを PCA で 2 次元へ射影し、テーマ(社会・政策/情報技術/災害・リスク)ごとに代表単語の配置と変化を示す。(表 2) 加えて、各年における単語の分野別ベクトルを結んだ三角形(分野間意味関係の幾何学表現)を用い、単語が「どの分野間で近接し、どの分野が乖離しているか」を幾何学的に把握する。Sankey 図(図 2)では、年ごとのラベル構成比と大局的な分布の安定性を示し、個別単語については PCA 射影図で具体的な文脈差を読む方針とする(本稿ではラベル遷移の細部列挙は省略する)。

4 結果

4.1 出現頻度推移クラスタの特徴

頻度推移の類型化により、分野固有の単語群と、外的刺激に応答して増減する単語群が整理された。建築構造分野では、材料・耐力・損傷等に関する専門単語が長期に安定して現れ、他分野とは異なる語彙領域が形成される傾向が見られる。一方、計画分野では災害・地域・居住に関わる単語の変動が比較的大きく、環境分野では省エネ・設備・空気質等、政策や技術動向に対応した単語群の増減が観測される。

代表例として、「避難」は大地震の翌年に計画分野で頻度が上昇する傾向を示し、災害が研究言説に短期的なピークをもたらすことを支持する。「感染」は 2020 年以降に急増し、環境分野だけでなく計画分野でも顕著な増加が見られる点で、分野横断的課題化を反映している。これらの頻度推移は、外的刺激が研究関心の量的配分を変えることを示すが、同時に「頻度が増えた単語が必ずしも同じ意味で使われているとは限らない」ため、次節の意味変化分析によって文脈差を検証する。

4.2 分野間意味関係ラベルの年次分布

分野間意味関係を幾何学的特徴に基づきラベル化したうえで、Sankey 図により年ごとの分布を可視化

表1 単語出現頻度の時系列推移に対するクラスタリング結果

	クラスタ1	クラスタ2	クラスタ3	クラスタ4
時系列特性※				
単語	試験, 結果, 図, 実験, 建築, 研究, 解析, モデル, 評価, 建物, 構造, 関係, 方向, 梁, 壁, 最大, 温度, 場合, 調査, 対象 (全1430語より抜粋)	変形, 耐力, せん断, 変位, 荷重, 応答, 剛性, 荷, 杭, 降伏, 応力, 破壊, 圧縮, 減衰, 加速度, 入力, 周期, 震, 塑性, ひび割れ (全489語より抜粋)	消費, 負荷, 空調, 熱, 電力, 風速, 湿度, 換気, 温, 気温, 暖房, 被験者, 日射, 断熱, 濃度, 冷房, 外気, 運転, 放射, 快適 (全409語より抜粋)	空間, 環境, 施設, 地域, 窓, 都市, 事例, 活動, 居住, 場所, 生活, 回答, 特徴, 分類, 照明, 内容, 地区, 人, 行動, 世帯 (全724語より抜粋)
例1				
例2				

— 計画 — 歴史・意匠 — 環境 — 構造

※時系列特性は、同一クラスタ内の全単語について出現頻度の年次推移を分野別に平均化し、分野間の出現頻度の規模差を考慮し、縦軸を対数スケールで表示した。

した。全体として、多くの単語は長期間同種の分野間関係を維持し、分野間の意味関係は大局的には安定していることが示される。これは、建築学研究において分野ごとの問題設定や語の使われ方が一定の構造を持ち、短期的な社会変化があっても言説全体が無秩序に再編されるわけではないことを示唆する。

一方で、特定年において分布が変動する局面も観測される。Sankey 図は、こうした年次変動を「全体の構造変化」として把握するための俯瞰図として位置づけられ、個別単語の変化の質（どの分野の文脈が変わったのか）については、次節のテーマ別 PCA 射影図で具体的に検討する。

4.3 テーマ別 PCA 射影による意味分岐

テーマ別の PCA 射影図は、単語が外的刺激に応答する際、分野ごとに異なる問題設定のもとで文脈が形成され、意味位置の分岐として現れることを示す。また、表 2 に入りきらなかった 2023 年の近傍語については、「感染」と「シェア」において重要な変化が見られたため別途表 3 に示す。

社会・政策（例：ZEB, シェア）では、政策語や

新技術語が特定分野で先鋭化し、他分野との距離が開くケースが見られる。「ZEB」は環境分野で専用的に用いられやすく、近傍語も環境系の専門文脈に集中するため、3 分野空間上で環境が相対的に離れやすい。これは、政策・制度によって導入された概念が、分野横断的に共有されるというより、まずは特定分野の技術語として定着する過程を反映している可能性がある。「シェア」は、同一表層語が時期によって抽象的意味（共有）から空間利用・運営の具体的意味へと傾くような変化が観察され、分野別配置の差として可視化される。

情報技術（例：ネットワーク, 自動）では、頻度が急増しない場合でも、近傍語の入れ替わりによって意味位置が緩やかに移動することがある。「ネットワーク」は人間関係や組織を指す語用から、構築・モデル化など情報技術的文脈へと傾くことで、分野間距離が再編される。これは、技術語が建築分野に浸透する際に、既存の一般語が専門的意味を獲得する過程を示す。

災害・リスク（例：避難, 感染）では、頻度の急増と同時に意味ベクトルの移動が顕著になりやすい。「感染」は、環境分野では換気・飛沫など空気質に関

表2 テーマ別にみた単語の意味変化推移分析結果

	社会・政策	情報技術	災害・リスク
例1	ZEB	自動	避難
近傍語 (注1)	歴史・意匠 近傍語なし 環境 なし→炭素-実現 →省エネルギー 計画 近傍語なし	歴史・意匠 災→販売→ベイ→瞬時 環境 変更→制御→制御→稼働 計画 洗浄→ドア→トイレ→運転	歴史・意匠 併記→経路→重視→給水 環境 災害→被災→災害→災害 計画 経路→震災→災害→経路
例2	シェア	ネットワーク	感染
近傍語	歴史・意匠 フィルム→浮上→睡眠→環境 環境 閲覧→事務所→遊び→病床 計画 共鳴→賃貸→集合住宅 →促進	歴史・意匠 直列→前述→暮らし→資源 環境 統合→分散→仮想→構築 計画 信頼→ワーカー→構築 →構築	歴史・意匠 なし×2→転居→平水 環境 汚物→飛沫→ウイルス →ウイルス 計画

注1) 本表では、代表年(2008, 2012, 2016, 2020)における最近傍語を時系列に示している。「近傍語なし」は、当該年における使用頻度が極めて低く、語の類似度を算出できなかったため、近傍語を検出できなかったことを示す。

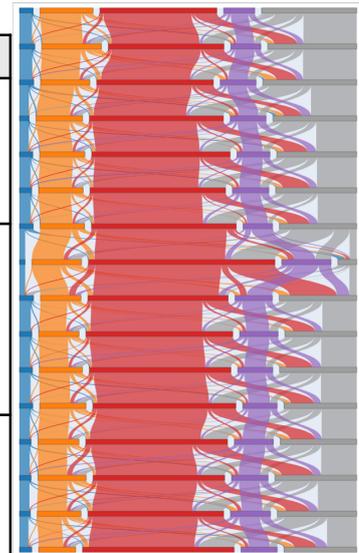


図2 単語の意味変化推移分析：ラベル間の時系列変化 (Sankey 図)。上から2008年, 下に向かって2023年を示す。ノードの色は, Divergent (青), Env-Design (オレンジ), Plan-Design (赤), Plan-Env (紫), Mixed (グレー)に対応する。

表3 「感染」「シェア」に対する2023年の近傍語上位20語

	感染	シェア
意匠・歴史	コロナ, ウイルス, 通風, 冬期下水, 聴覚, 対策, ピーク, 上水, 高温, 小型, 太平洋, エンド, 沈下, 度, バリアフリー, 熱中, 入射, ヒートアイランドモジュール	輸送, 需要, 鉄道, 終了, 回復, 増加, 生産, 高度, 減少, 空き家, 低下, 病床, テナント, 大幅, 卓越, 工場, 戸数, 序, 搬送, 列車
環境	ウイルス, リスク, コロナ, エアロゾル, 換気, 濃度, 飛沫, 汚染, 粒子, 物質, 発生, 対策, 排気, 空気, 気流, 清浄, 室内, 曝露, 呼吸	テーブル, ベッド, 棚, シルト, 寝室, 浴室, 家具, 間隙, 洗面, トイレ, 市街, 歪み, リビング, 洋室, 比熱, 重度, 揺れ, キッチン, レンジ, 過小
計画	コロナ, ウイルス, 児童, 生徒, 学校, 対応, 対策, 拡大, ケア, 教室, 室, 授業, 職員, 保育, 制限, 配慮, 確保, 子ども, 自由, 保護	カフェ, 開催, 用意, 半分, 強化, 活性, 同居, 飲食, 町内, 商店, 共同, 安定, 新た, テナント, まとめ, ほか, 株式会社, まち, つながり, 宿泊

わる語が近傍語として現れ、計画分野では学校・保育など集団生活の場に関わる語が現れるなど、同一事象に対しても分野ごとに異なる課題設定が前景化する。この分岐は、外的刺激が単語の意味を一様に変えるのではなく、分野固有の研究対象・方法・評価軸によって意味変化が方向づけられることを示している。

ている。

5 まとめ

本研究は、2008-2023年の日本建築学会大会梗概集を対象に、頻度推移と意味変化を統合して建築学研究言説の再編過程を定量化した。頻度推移は外的刺激への応答(関心の量的配分)を捉える一方、意味ベクトルは分野ごとの問題設定の差(関心の質的差異)を可視化する。

Sankey図は分野間意味関係の大局的安定性と年次変動を俯瞰する役割を担い、テーマ別PCA射影図は、外的刺激に応答する単語が分野ごとに異なる文脈を形成し、意味位置の分岐として現れることを具体的に示す。今後は、サブ分野単位での細分化や、制度改正・技術標準等の外的要因とのより厳密な対応付け、設計実務文書への拡張によって、知識再編の機序をさらに詳細に明らかにする。

参考文献

- [1] 日本建築学会. 大会梗概集(建築構造 I, II, III, IV, 環境工学, 建築計画, 建築歴史・意匠), 2008-2023.

<https://www.aij.or.jp/taikaidvd.html>.

- [2] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. In Katrin Erk and Noah A. Smith, editors, **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1489–1501, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [3] 相田太一, 小町守, 小木曾智信, 高村大也, 持橋大地. 異なる時期での意味の違いを捉える単語分散表現の結合学習. 自然言語処理, Vol. 30, No. 2, pp. 275–303, 2023.
- [4] Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. SemEval-2020 task 1: Unsupervised lexical semantic change detection. In Aurelie Herbelot, Xiaodan Zhu, Alexis Palmer, Nathan Schneider, Jonathan May, and Ekaterina Shutova, editors, **Proceedings of the Fourteenth Workshop on Semantic Evaluation**, pp. 1–23, Barcelona (online), December 2020. International Committee for Computational Linguistics.