LLM 支援アノテーションによる『ユリシーズ』意識モードの定量化と転調点分析

— evidence-anchored prompting と半教師あり GMM による探索的検証 —

（研究報告 / draft）

要旨（Abstract）

本稿は、J. Joyce『Ulysses』の 555 thought-span（場面単位）に対し、大規模言語モデル（LLM）を用いた証拠語句抽出つきアノテーションを実施し、得られた特徴ベクトルを半教師あり混合ガウスモデル（GMM）で潜在状態に写像した。事後確率のエントロピーと上位 2 状態の確率差（margin）により転調点を定義し、内言・知覚・推論等の拮抗が生じる箇所を上位 10 例として抽出した。証拠語句に Attention を固定するプロンプト設計により判定の揺らぎと幻覚が抑制され、低コストで解釈可能な注釈・可視化が可能であることを示した。

1. はじめに（Introduction）

『ユリシーズ』は意識の流れ（stream of consciousness）を極限まで押し進めた作品であり、語りの位相（内言／対話／知覚／推論など）が短いスパンで切り替わる。この切替は読解の核心に関わる一方、網羅的な注釈づけは人的コストが高い。近年のデジタル人文学は、テキスト分析・アーカイブ・可視化を通じて解釈支援を目指してきたが、細粒度の解釈ラベル付与は依然としてボトルネックである。

1.1 関連研究

デジタル人文学における注釈・コメントリーの整備は、作品を研究資源として再利用可能にする基盤である。『ユリシーズ』についても、学術的デジタル版（digital scholarly edition）を構想し、多版・草稿・文脈資料・注釈をネットワーク化する試みが議論されている[1]。一方、近年の LLM はテキスト分類・要約・コーディング作業の補助に用いられ、人手アノテーションと同程度の一致やコスト削減を報告する研究が増えている[2][3]。ただし、LLM 出力はプロンプトや確率的生成により揺らぎうるため、再現性・説明可能性が課題となる。推論過程（chain-of-thought）を明示させる、あるいは複数推論の整合を取る研究は、品質向上の実践的知見を与えている[4][5]。

1.2 本研究の目的と貢献

本研究の目的は、(i) LLM 支援により『ユリシーズ』の細粒度アノテーションを低コストに実施し、(ii) その特徴量から潜在状態（意識モード）を推定し、(iii) 曖昧さの高い箇所を転調点として抽出・解

釈する枠組みを提示することである。貢献は、証拠語句抽出（evidence-anchored prompting）を核に、判定の安定性と解釈可能性を同時に確保した点にある。

2. 手法（Methods）

2.1 データ：thought-span の定義

分析単位は、本文を連続する短区間に分割した thought-span（場面）である。各 span は原文（英語）と対応する日本語要約・訳注を付与し、全 555 span を対象とした。区切りは、話者交替・段落境界・顕著な主題転換など、読解上のまとまりを基準として人手で定義した。

2.2 LLM 支援アノテーションと証拠語句抽出

各 thought-span に対し、意識モード（例：inner_speech, perception, reasoning, dialogue, emotion, memory, imagination, quotation など）を中心とする複数ラベルを LLM に付与させた。重要点は、分類ラベルに加えて「根拠となる原文の 3〜12 語（evidence）」を抽出させ、その語句に基づいて判定理由（短い説明）を生成させたことである。以後の数値化・検証は、抽出された evidence を主参照として実施した。

2.3 再現性のためのプロンプト設計（アンカー効果と幻覚抑制）

LLM に自由回答で「意識モードは何か」と問うと、広い文脈からの推論が働き、出力が揺らぎやすい。これに対し、本研究ではまず evidence を抽出させ、注意（Attention）を特定の語句列へ固定する。この操作は探索空間を狭めるアンカーとして機能し、その後のラベル付与と理由づけが同一根拠に収束しやすくなる。また「原文にない断定をしない」「根拠語句を必ず提示する」といった制約を設けることで、引用不可能な情報に基づく幻覚（hallucination）が抑制される。実装上は温度（temperature）を低く設定し、同一プロンプト・同一入力での再実行により安定性を点検した。

2.4 特徴量設計とベクトル化

各 span を特徴ベクトル $x_t ¥in ¥mathbb{R}^d$ に写像した。特徴量は、(1)意識モードの確率分布、(2)place/style/myth など解釈上重要な補助次元、および(3)転換タイプ（contrast, intrusion, rumination, semantic_association, script_switch 等）から構成した。place/style/myth は、場所参照・文体操作・神話的連関の強さを 0〜1 のスコアとして定義し、LLM の根拠語句と要約を参照して付与した。

2.5 半教師あり混合ガウスモデル（GMM）

潜在状態 $z_t ¥in ¥{1,¥dots,K¥}$ を仮定し、観測 x_t が K 個のガウス分布の混合から生成されるとする。混合比を $¥pi_k$、平均 $¥mu_k$、共分散 $¥Sigma_k$ とすると、尤度は $$p(x_t)=¥sum_{k=1}^K ¥pi_k¥,¥mathcal{N}(x_t¥mid ¥mu_k,¥Sigma_k)$$で与えられる。学習後、各 span の事後確率 $p(z_t=k¥mid x_t)$ を算出し、状態割当（top1）と確率（top1_p）を得た。

2.6 転調点（Turning Point）の定義：エントロピーと margin

状態が一意に定まらない曖昧さを定量化するため、事後確率のエントロピー H_t を定義した。$$H_t=-¥sum_{k=1}^{K} p(z_t=k¥mid x_t)¥,¥log p(z_t=k¥mid x_t)$$また、上位 2 状態の確率差を $¥mathrm{margin}_t = p_{(1)}-p_{(2)}$（$p_{(1)}¥ge p_{(2)}$）とし、小さいほど拮抗が強いと解釈した。本研究では H_t の高い順に候補を抽出し、evidence と本文を参照して転調点として解釈した。

2.5 再現性と公開実装

本研究の手順は、公開された Jupyter Notebook として実装されており、研究者・学生を含む第三者が同一の手続を追試できる形で提供される。再現性の確保は、(i) 分割単位（thought-span）の明示、(ii) アノテーション設計の固定、(iii) 特徴量計算とモデル推定のスクリプト化、(iv) 生成物（CSV/図）の保存、の 4 点により担保した。

再現手順の概要を以下に示す。第一に、GitHub 上で公開したリポジトリ（Mokafe-ulysses-narrative-gmm）を取得し、notebooks/00_ulysses_gmm_report_v1_02.ipynb を実行する。第二に、必要な Python 依存関係を準備した上で、(a) thought-span 抽出、(b) LLM による根拠語句（evidence）の抽出とアノテーション、(c) 特徴量ベクトル化、(d) GMM 推定と事後確率の算出、(e) 境界性（entropy/margin）に基づく転調点抽出、を順に実行する。第三に、実行後に boundary_report.csv、turning_points_top10.csv 等の中間生成物が保存され、図表と表（Top10 転調点）を含む結果が得られる。

誰でも試せることを重視し、実行環境は一般的な Python/Jupyter 構成を前提とする。ローカル環境に加え、Google Colab 等のクラウド実行環境でも Notebook 単体で再現できるようにし、入力データ（対象テキスト、章区切り、設定ファイル）と出力ファイル名を固定した。

なお、LLM の推論は確率的であり出力が揺らぎ得るが、本研究では「根拠となる短い語句列（3〜12 語）を原文から抽出せよ」という制約を付与し、その語句列を後続の数値化の入力として保存することで、判断の探索空間を限定した。これにより、雰囲気的な解釈に依存しない計算経路が形成され、再実行時の変動が小さくなる傾向が観察された。実務運用では、temperature=0 等の決定論的設定と併用することで、さらに安定性を高められる。

公開実装により、(i) 別作品・別章への横展開、(ii) 特徴量設計の差し替え（例：語彙難度、修辞、引用形態の追加）、(iii) 状態数 K や初期値の感度分析、が低コストで実施できる。したがって本枠組みは、文学研究の解釈補助ツールとして再利用可能である。

（参照）GitHub Notebook: https://github.com/Mokafe/Mokafe-ulysses-narrative-gmm/blob/main/notebooks/00_ulysses_gmm_report_v1_02.ipynb

3. 結果（Results）

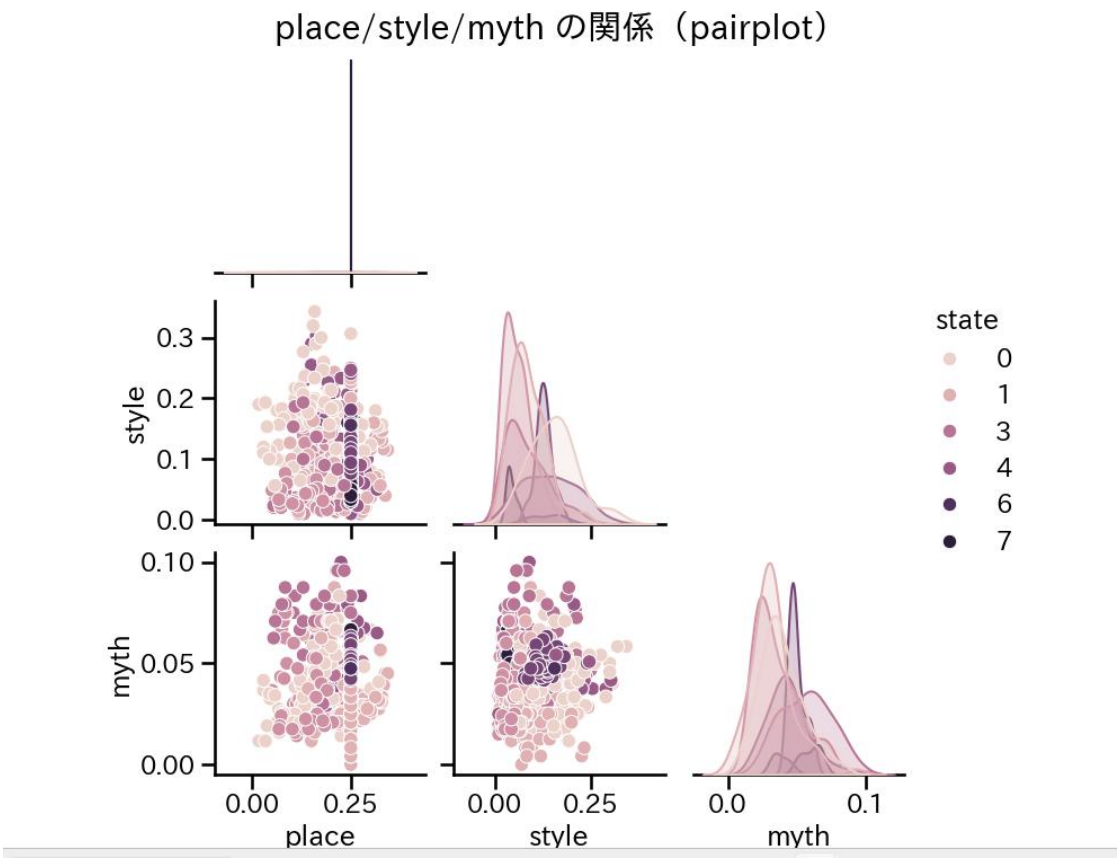
3.1 潜在状態数 K と状態解釈（探索的）

本実験では $K=8$ を採用した（state は 1〜8）。各状態は、アノテーション特徴の重心と代表 span の読解に基づき、内言優勢・知覚優勢・推論優勢・対話優勢などの解釈的ラベルを付与した。以後、表記は state（GMM 推定）と上位状態（top1_state）を区別する。

3.2 place/style/myth の相関と状態分布

place/style/myth の 3 次元について、状態ごとの分布を図 1 に示す。視覚的には、style の高い領域に特定状態が集中する傾向が見られ、myth は右裾が重い分布として現れた。このような補助次元は、状態の解釈（例：文体操作が強い場面の集合）を支える。

図 1　place/style/myth の関係（pairplot）。状態（state）ごとの分布を示す。



3.3 転調点（Turning points）上位 10 例

エントロピー H_t の高い順に抽出した転調点上位 10 例を表 1 に示す。上位では top1 と top2 の確率が拮抗し、margin が小さい例が含まれる（例：chapter1 span7）。

rank	chapter	span_id	transition_type	entropy	top1_p	top2_p	evidence_best
172	1	7	rumination	0.693457	0.502376	0.497594	疑問文＋「Thumping」反復で作業音が

							思考に食い込む。
280	2	11	semantic_association	0.685999	0.559711	0.440289	視覚（eye）＋読字（read）。
173	3	7	semantic_association	0.554764	0.837654	0.091454	断片文が連なり、内言の流れが優位。
292	4	11	semantic_association	0.536416	0.772354	0.227646	By の反復で列挙リズムが支配。
502	5	17	semantic_association	0.534026	0.774302	0.225698	聴覚入力（sound, chime）でアンカーが補強。
31	6	2	intrusion	0.399937	0.876068	0.117514	I hear から連鎖する感覚-像の急激な挿入。
346	7	13	script_switch	0.391869	0.867105	0.132894	固有名の提示と姿勢描写により script_switch と social/emotion ...
178	8	7	semantic_association	0.362360	0.882283	0.117716	「Reads」「Quickly」「practice」で観察→推測が同一流れ。
126	9	5	rumination	0.349407	0.888612	0.111388	「Poor papa!」の感嘆から過去行為（didn’t go）へ移る。
514	10	18	contrast	0.338288	0.893897	0.106103	「polite」＋対人対象の列挙。

表 1　転調点（Turning points）Top10（entropy 順）。

3.4 転調点の詳細分析（例）

例 1：chapter1 span7（rank172, H=0.693, margin=0.0048）。ここでは「But will he save the circulation? Thumping. Thumping.」のように疑問文と作業音の反復が現れ、内言（inner_speech）と推論（reasoning）がほぼ同確率で拮抗した。外部刺激（反復音）が思考へ割り込むことで、自己内対話が断続的に中断される瞬間を、確率の均衡として捉えている。

例 2：chapter2 span11（rank280, H=0.686, margin=0.119）。「Bloowhose dark eye read Aaron Figatner's name.」のような視線・読字の描写が根拠語句として抽出され、知覚（perception）と内言（inner_speech）の間で状態が揺れた。視覚行為が単なる描写に留まらず、固有名の読取りを通じて内的連想へ移行する端点として解釈できる。

例 3：chapter5 span17（rank502, H=0.534, margin=0.549）。「The sound of the peal of the hour of the night...」に代表される聴覚入力がアンカーとなり、知覚（perception）と推論（reasoning）が競合した。時刻を告げる鐘の音が、場面の空間配置（place）と意味づけ（reasoning）を同時に呼び起こす点で、“知覚—解釈”の境界として位置づけられる。

4. 考察（Discussion）

4.1 低コスト注釈と解釈支援としての意義

本枠組みは、専門家が長期間かけて行ってきた注釈作業の一部を、LLM 支援により短時間・低コストで再現可能にする。重要なのは、単なる自動分類ではなく、根拠語句（evidence）と理由づけを同時に残す点であり、後続の読解・批評における検証可能性を担保する。さらに、状態推定と曖昧さ指標により、読解上の“引っかかり”を候補として提示できる。

4.2 安定性・妥当性に関する留意点

第一に、LLM は確率的生成であるため、同一入力でも完全一致を保証しない。本研究では evidence 抽出によるアンカー効果、低温度設定、および「原文に根拠のない断定を避ける」制約で揺らぎを抑制したが、再現性評価（複数回実行の一致率、温度感度、モデル差）を体系的に行う必要がある。第二に、特徴量設計は解釈枠組みを事前に規定する。特に place/style/myth の定義は研究者の問題設定を反映するため、相互評価（複数注釈者・専門家レビュー）による妥当性検証が望ましい。

4.3 今後の展開

今後は、(i) 一部区間を専門家が再注釈し、LLM 注釈との一致・差分を分析すること、(ii) 転調点候補の提示を読解インタフェースに統合し、読者の解釈プロセスに与える影響を評価すること、(iii) アクティブラーニング（曖昧例を優先的に再注釈）により、最小コストで注釈品質を上げる運用を検討することが挙げられる。

5. 結論（Conclusion）

本稿は、『ユリシーズ』の thought-span を対象に、証拠語句抽出つき LLM アノテーションを特徴量化し、半教師あり GMM で潜在状態を推定した。事後確率エントロピーと margin に基づく転調点抽出により、内言・知覚・推論などが拮抗する箇所を体系的に提示できた。evidence-anchored prompting は、判定の揺らぎと幻覚を抑制しつつ解釈可能な痕跡を残す点で有効である。

参考文献

[1] "Networking Ulysses"（digital scholarly edition 構想に関する解説）

[2] Gilardi, F., Alizadeh, M., Kubli, M. (2023). ChatGPT outperforms crowd-workers for text-annotation tasks. PNAS.

[3] He, X. et al. (2024). Large language models are human-level annotators. Nature Machine Intelligence.

[4] Wei, J. et al. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. arXiv:2201.11903.

[5] Wang, X. et al. (2022). Self-Consistency Improves Chain of Thought Reasoning in Language Models. arXiv:2203.11171.

[6] （参考）銀河鉄道の夜を対象とした boundaryness/転調点分析の先行研究（添付 PDF）。