読者の質問と予想が駆動する物語生成

城戸 晴輝 [1]　上垣外 英剛 [2,3]　村上 聡一朗 [2]　高村 大也 [1]　奥村 学 [1]
[1] 東京科学大学　[2] 株式会社サイバーエージェント　[3] 奈良先端科学技術大学院大学
{haruki,takamura,oku}@lr.pi.titech.ac.jp
kamigaito.h@is.naist.jp　murakami_soichiro@cyberagent.co.jp

概要

近年の大規模言語モデルの発展に伴い物語生成技術は大きく進歩しているものの，既存研究は作者視点に偏り，読者の質問や予想といった認知プロセスを十分に考慮していない．本研究は，物語完結感と認知的興味の理論を統合し，読者エージェントの予想を裏切りつつ事後的に説明可能な質問の解決をプロットに組み込む手法を提案する．これにより完結感だけでなく，意外性と納得感による認知的興味の両立を目指した．実験の結果，提案手法はベースラインと比較して創造性や矛盾のない物語構造だけでなく，複雑さや関心など多くの側面で高い性能を示すことがわかり，文学理論の知見を生成プロセスに組み込むことと，読者視点を導入するアプローチの有効性を示唆した．

1　はじめに

近年の大規模言語モデル (LLM) の発展により，数千単語に及ぶ物語の生成が現実的なタスクとなりつつあるものの，既存の物語生成手法の多くは，プロット構築や執筆といった作者の作業をモデル化することに主眼を置いており「読者が物語をどう解釈し，何を期待して読むのか」といった認知プロセスを十分に考慮していない．その結果，生成された物語は局所的には整合していても，全体として平坦で予測可能な展開に終始する場合がある．

文学理論の観点からは，物語の面白さは読者の認知的な働きかけに依存するとされる．Carroll は，物語の進行は読者が抱く質問の発生と解決によって駆動される [1] とし，Kintsch は，読者の予想を裏切りつつも事後的には納得できる展開こそが認知的な興味を喚起すると論じている [2]．すなわち，質の高い物語を生成するためには，作者視点だけでなく，読者の質問と予想をシミュレートし，それを巧みに操作する機構が必要である．そこで本研究では，1000 単語程度の物語生成において，文学理論に基づく読者モデルを組み込んだ新たな生成フレームワークを提案する．本手法では，読者エージェントが生成過程の物語に対して抱いた質問と予想 (表 1) を意図的に裏切ることで物語を展開させる．本研究では，読者の認知プロセスを明示的に扱うことが物語や読書体験の質にどう寄与するかを検証する．

表 1　質問とその予想の例

Question: Why did Alex get the new haircut? **Expectations:** To express or affirm a change in identity. / Purely aesthetic, just wanted a new look ...

2　関連研究

2.1　文学理論

物語の構造や面白さを定義する上で，読者の認知プロセスに着目した理論が数多く提案されている．Carroll は Narrative Closure という概念を提案し，「物語の完結は，物語の大部分の問いであるマクロ質問と，それに従属する全てのミクロ質問が解決されることにより生じる」と定義している [1]．すなわち，読者はそれらの提示された質問が構造的に全て解消されたときに，物語が終わったという感覚を得るのである．しかし，単に質問が解決されるだけでは不十分である．Kintsch は，読者の認知的興味を喚起する物語の条件として，「予測不可能であり，かつ事後説明可能である」 ことを挙げている [2]．彼によれば，完全に予測可能な展開には興味が生じない．一方で，予測からの逸脱が興味をもたらすには，事後的にはその理由や意味が説明可能（事後説明可能）であることが必須条件であると論じている．本研究は，この 2 つの理論を統合することで読者の認知的興味を高く保ちつつ，確かな完結感を与えるプロセスをモデル化するものである．

2.2 物語生成タスク

読者の質問を物語生成に活用する試みは LLM の登場以前から存在する．Bailey は，物語生成を読者が抱く質問や期待によって生じる物語性の探索プロセスとして定義し，読者視点を取り入れたモデルの構想を提案している [3]．近年では LLM を活用した物語生成が盛んに行われている．Huot らは執筆プロセスを分業化する Agents' Room を提案し，高品質な生成を実現した．彼らは生成手法に加え，物語の一貫性や創造性を測るための評価プロンプトも提案している [4]．読者の読書体験の評価に特化した研究として，Harel-Canada らは PDS (Psychological Depth Scale) を提案し，読者の読書体験を多角的に評価する包括的な指標の重要性を示している [5]．物語構造の制御に関しては，Brei らが Narrative Closure を応用し，始点と終点を事前に関連付けることで完結感を保証する手法である RENARGEN を提案した [6]．また Castricato らは，物語生成を因果関係に関する質問応答の連鎖として捉えることで一貫性を保つ手法を提案している [7]．これら既存の LLM ベースの手法は，静的な構造決定や整合性に主眼を置くものである．しかし，LLM を用いたエージェントにより読者の認知プロセスを明示的にシミュレートし，文学理論の知見を工学的に具体化する試みは未だ行われていない．

3 提案手法

本研究では，Kintsch の認知的興味および Carroll の物語完結感の理論を工学的に実装するため，LLM エージェントによる物語生成の先行手法である Agents' Room [4] を拡張し，マルチエージェントシステムによる読者の質問と予想が駆動する物語生成手法を提案する．本手法の核心は，読者の質問を解決へ導くプロット構築を主軸とし，その解決方法として予測不可能であり事後説明可能な展開を生成する点にある．これにより物語完結感と認知的興味の達成を実現する．

3.1 エージェント構成

本システムは，Agents' Room [4] に対し，読者の認知プロセスを模倣する読者エージェントと，物語を牽引する質問を管理する質問管理エージェントを新たに導入している．各エージェントの役割は以下の通りである．

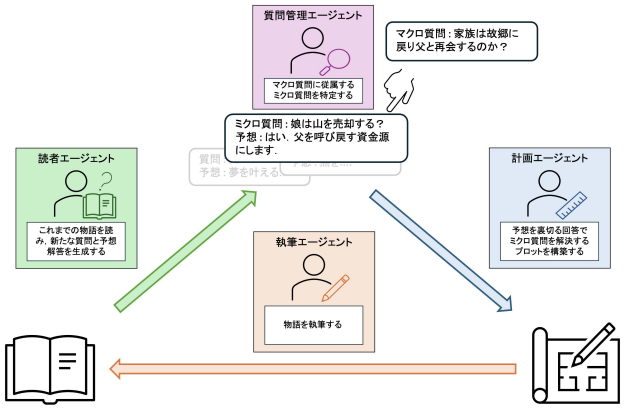

図 1 提案手法の図解（中盤生成フェーズ）

読者エージェント

- **質問の生成:** 物語の因果関係に基づいて自然に生じる質問と回答の予想を生成する．
- **事後説明可能性の評価:** 予想外に解決されたミクロ質問について納得感を評価する．

質問管理エージェント

- **マクロ質問の定義:** 物語における主要な問いであるマクロ質問の計画を行う．
- **ミクロ質問の特定:** 読者から発生した質問からマクロ質問に従属するミクロ質問を特定する．

計画エージェント

- **基本設定 (葛藤・舞台・人物) の計画:** 物語全体における基本設定を計画する．
- **プロット構築:** 予測不可能かつ事後説明可能に質問解決をするプロットを構築する．

執筆エージェント

- **物語の執筆:** 物語の設定やプロットを元に物語を執筆する．

3.2 生成アルゴリズム

本手法では，物語構成として Freytag のピラミッド [8] を採用し，物語を 5 つのセクションに分割し順次生成する．またプロット構築における質問解決の戦略の違いに基づき，これらを以下の 3 つのフェーズに分類して制御を行う．

- **序盤 (Exposition):** 物語の基盤構築に注力する．
- **中盤 (Rising Action, Climax, Falling Action):** 予測不可能性を保証する．
- **終盤 (Resolution):** 事後説明可能性を保証する．

表 2　提案手法のバリエーションごとのプロット構築プロンプトにおける質問回答指示部分の違い

バリエーション	対象	プロンプトへの指示内容
Closure	全体	*Dramatize the answer to the following **[micro/macro]-question**.* *Ensure that you also generate other plot points necessary to bridge these events and advance the narrative naturally. Do not move straight from question to answer; make the resolution feel earned and story-shaped by adding obstacles, reversals, detours, and meaningful choices with consequences.*
+Unpredictable	中盤	*+ However, answer in an **unpredictable** way that deviates from the reader's expectations.*
+Postdictable	終盤	*+ However, answer in a **postdictable** way that is well-motivated by the already answered micro-question.*

3.2.1 初期計画フェーズ　まず，計画エージェントが入力プロンプトに基づいて物語の基本設定（葛藤・人物・舞台）を計画する．続いてこれらの設定に基づき質問管理エージェントが物語における主要な問いであるマクロ質問を定義する．

3.2.2 序盤生成フェーズ　物語の序盤では，計画エージェントが初期設定に基づいたプロットを構築し，執筆エージェントが物語を執筆する．ここでは質問や予想に基づく制御は行わず，物語の基盤構築に注力する．

3.2.3 中盤生成フェーズ　物語の中盤では，図 1 に示す以下のサイクルを実行する．まず，読者エージェントがこれまでの物語を読み，新たな質問と予想解答を生成する．その中から質問管理エージェントがマクロ質問に従属するミクロ質問を一つ特定する．次に計画エージェントが予想を裏切る回答でミクロ質問を解決する続きのプロットを構築し，執筆エージェントが続きのセクションの物語を執筆する．このプロセスにより予測不可能性を保証する．

3.2.4 終盤生成フェーズ　物語の終盤では，読者エージェントがこれまでの物語を読み，ミクロ質問のうち予想外に解決され未だ納得できていない質問を一つ特定する．次に，計画エージェントがそのミクロ質問に動機づけられた回答でマクロ質問を解決するプロットを構築し，執筆エージェントが物語を執筆する．このプロセスによりミクロ質問を物語全体の中に統合し，事後説明可能性を保証する．

4　実験

4.1　実験設定

生成モデルには `gemini-3-flash-preview` と `gpt-5-mini` を使用した．データセットには，Tell Me a Story [4] のテストデータ 55 件を使用し，1000 単語の物語を生成した．各データにつき 3 回の生成試行を行った．詳細は付録 A に記載する．

4.2　比較手法

提案手法の性能を評価するため，以下のベースラインモデルを設定した．

End-to-End (E2E):　Tell Me a Story のプロンプトをもとに一括して物語を生成する手法．

Agents' Room (AR):　Huot ら [4] が提案した既存手法．プロット構築は物語全体の開始前に一度だけ行われる．各セクション 200 単語を目標単語数とし，合計 1000 単語の物語を執筆する．

Agents' Room Step (AR-Step):　AR のプロット構築をセクションごとに行うように変更した手法．

AR-Step (Unpredictable):　AR-Step のプロット構築時に読者のプロット予想を裏切ることを明示する手法．

提案手法としては，ミクロ/マクロ質問の解決・予測不可能性・事後説明可能性の効果を検証するため，表 2 に示す 3 つのバリエーションを用意した．

Ours (Closure):　AR-Step のプロット構築時にミクロ質問とマクロ質問の解決を明示する手法．

Ours (+Unpredictable):　読者の予想を裏切る回答でミクロ質問を解決する手法．

Ours (+Unpredictable+Postdictable):　読者の予想を裏切る回答でミクロ質問を解決し，予想外で未だ納得できないミクロ質問に動機づけられた回答でマクロ質問を解決する手法．生成例を付録 B に記載する．

4.3　評価方法

評価には LLM-as-a-judge [9] によるペアワイズ比較を採用した．評価モデルは生成モデルと同じモデルを使用し，位置バイアスを軽減するため，提示順序を入れ替えた 2 回の評価を行った．得られた勝敗データからのスコア算出には，Bradley-Terry モデル [10] を採用した．`gpt-5-mini` による実験結果は付録 C に記載する．評価指標には以下の 2 種類を用い

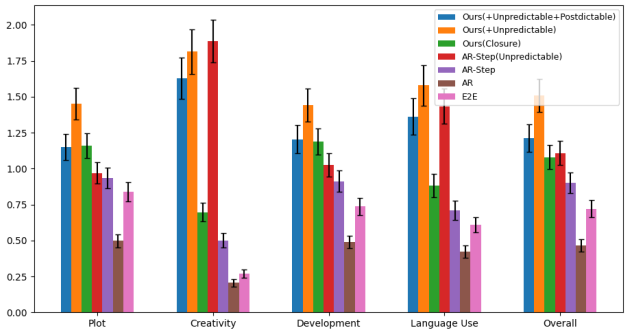

図 2　*AR Metrics* での評価結果（gemini-3-flash-preview）

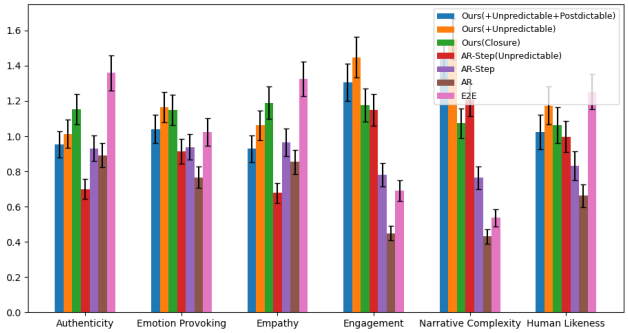

図 3　*PDS* での評価結果（gemini-3-flash-preview）

た．評価指標の詳細は付録 D に記載する．

***Agents' Room Metrics*:**　Plot, Creativity, Development, Language Use, Overall の 5 つの尺度から構成され，物語としての一般的な品質を評価する指標 [4] である．読者の質問や予想に基づいて物語の構成や展開を制御する影響を評価するため，物語の構成 (Plot)，設定の具体化 (Development)，展開の意外性 (Creativity) の 3 点に特に着目する．

***Psychological Depth Scale*:**　Authenticity, Emotion Provoking, Empathy, Engagement, Narrative Complexity, Human Likeness の 6 つの尺度から構成され，読者の読書体験の質を評価する指標 [5] である．認知的興味への影響を評価するため，解釈の再構築を促す物語構造の複雑さ (Narrative Complexity) と，維持される関心の強さ (Engagement) の 2 点に特に着目する．

4.4　結果

Agents' Room Metrics　*AR Metrics* における評価結果を図 2 に示す．ベースラインと比較し，提案手法 Ours (+Unpredictable) が Overall において高い性能を示すことから，読者の認知プロセスを明示的に扱う本手法は物語の全体的な品質を高める上で有効であることが示された．また，Creativity において AR-Step (Unpredictable) も高い性能を示すことから，読者モデルの予想を裏切ることで Creativity を高めることが確認された．さらに，質問を深掘りするプロセスにより登場人物の行動動機などが詳細化され Development が向上した．また，ミクロ/マクロ質問の解決により物語を確実に収束させたことに加え，予測不可能性の導入により転換点となるイベントが生成されたことで Plot が向上した．なお，Ours (+Unpredictable) と比較して Ours (+Unpredictable+Postdictable) にさらなる性能向上は確認されず，事後説明可能性を明示的に保証したことによる定量的な改善は見られなかった．原因として，過去に基づいた解決指示により物語の整合性や展開の自然さが損なわれた可能性が示唆される．詳細は付録 E に記載する．

Psychological Depth Scale　*PDS* における評価結果を図 3 に示す．ベースラインと比較し，提案手法 Ours (+Unpredictable) が Narrative Complexity と Engagement においても高い性能を示すことから，読者の認知プロセスを明示的に扱う本手法は読者の認知的な読書体験を高める上でも有効であることが示された．特に，*AR Metrics* の Creativity では AR-Step (Unpredictable) との間に顕著な差は見られなかったものの，Narrative Complexity において高い性能を示した．これは質問回答のプロセスがもたらす強固な因果関係が構造的な深みとして機能した結果である．このように，文学理論の知見をモデルに組み込むことで，更なる改善が見られることがわかる．Ours (+Unpredictable+Postdictable) については，*PDS* においても Ours (+Unpredictable) と比較してさらなる性能向上は確認できなかった．

5　おわりに

本研究では，Carroll の物語完結感と Kintsch の認知的興味の理論に基づき，読者の質問と予想が駆動する物語生成手法を提案した．実験の結果，読者の認知プロセスを明示的に扱う本手法は，物語の全体的な品質を高めるだけでなく，認知的な読書体験を高める上でも有効であることが示された．また，読者モデルの予想を裏切ることで Creativity を高めることが確認された．今後の展望は，認知的興味だけでなく，情緒的興味を両立した物語生成を実現することである．

参考文献

[1] Noël Carroll. Narrative closure. **Philosophical Studies**, Vol. 135, No. 1, pp. 1–15, 2007.

[2] Walter Kintsch. Learning from text, levels of comprehension, or: Why anyone would read a story anyway. **Poetics**, Vol. 9, No. 1-3, pp. 87–98, 1980.

[3] Paul J. Bailey. Searching for storiness: Story-generation from a reader's perspective. 1999.

[4] Fantine Huot, Reinald Kim Amplayo, Jennimaria Palomaki, Alice Shoshana Jakobovits, Elizabeth Clark, and Mirella Lapata. Agents' room: Narrative generation through multi-step collaboration. In **The Thirteenth International Conference on Learning Representations**, 2025.

[5] Fabrice Y Harel-Canada, Hanyu Zhou, Sreya Muppalla, Zeynep Senahan Yildiz, Miryung Kim, Amit Sahai, and Nanyun Peng. Measuring psychological depth in language models. In **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing**, pp. 17162–17196, 2024.

[6] Anneliese Brei, Chao Zhao, and Snigdha Chaturvedi. Returning to the start: Generating narratives with related endpoints. In **Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)**, pp. 101–112, 2024.

[7] Louis Castricato, Spencer Frazier, Jonathan Balloch, and Mark Riedl. Tell me a story like i'm five: Story generation via question answering. **Proceedings of the 3rd Workshop on Narrative Understanding**, 2021.

[8] Gustav Freytag. **Freytag's Technique of the Drama: An Exposition of Dramatic Composition and Art**. 1896.

[9] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. Judging LLM-as-a-judge with MT-bench and chatbot arena. In **Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track**, 2023.

[10] Ralph Allan Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. **Biometrika**, Vol. 39, No. 3/4, pp. 324–345, 1952.

Prompt	Write a story about a Greek mythological character meeting someone from the present in a cafe. ... The narrator recognizes the mythological character for what she is.
Story	Elias, a chronic wanderer, sat in a warm cafe fighting the urge to run back to the cold, lonely road. When the Goddess Hestia appeared, she brought a silent challenge: stay or flee. As the biting wind of his past battered the door, Elias chose to fight, locking the world out to protect the hearth. Hestia blessed him as the true keeper of the flame, and he burned his travel bag in the fire. No longer a ghost, Elias accepted the cafe's key and finally stopped running.

表 3　プロンプトと生成された物語（要約）の例

A　実験設定の詳細

生成条件の統制について，以下の 2 点を厳密に設定した．第一に，出力長による評価への影響を排除するため，目標単語数を W とした際，生成される単語数が $0.9W$ から $1.1W$ の範囲に収まらなければ再生成を行うという制約を設けた．第二に，生成の分散を抑え，純粋なアルゴリズムの差異を比較するため，初期計画フェーズ（葛藤・人物・舞台）の制御を行った．具体的には，3 回の試行それぞれにおいて個別の初期計画を生成し，同一試行内においては全ての手法で共通する計画を使用した．これにより，各試行で物語の前提条件を揃えつつ，試行間では異なる設定を用いることで，多様なシナリオに対する頑健な比較を可能にした．

B　生成例

提案手法 Ours (+Unpredictable+Postdictable) について `gemini-3-flash-preview` で生成・要約したものを表 3 に記載する．

C　`gpt-5-mini` を用いた実験

生成モデルと評価モデルに `gpt-5-mini` を用いた実験の結果を図 4，5 に示す．本文の `gemini-3-flash-preview` を使った場合と比べ，AR-Step と提案手法間の性能順に関してはほとんど同じであるが，E2E 性能が低く評価されている．また，多くの側面で提案手法群よりも AR-Step (Unpredictable) の性能が高い．これは，質問回答指示により展開の自由度が下がる影響が顕著に現れたためと考えられる．

D　評価指標の詳細

Agents' Room Metrics　以下の 5 つの側面について評価を行う [4].

- Plot: 明確な構成（導入・中盤・結末）を持ち，出来事や転換によって矛盾なく物語が進行しているかを評価する．
- Creativity: ステレオタイプを避け，独自の要素や魅力的なテーマを含んでいるかを評価する．
- Development: 登場人物や設定に十分な詳細と背景があり，物語に説得力があるかを評価する．
- Language Use: 語彙や文構造が多様で，修辞技法を用いて単調な言い回しを回避しているかを評価する．
- Overall

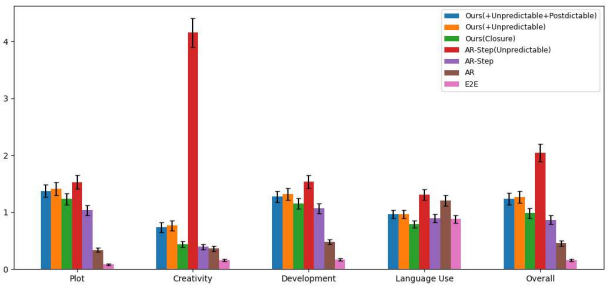

図 4　*AR Metrics* での評価結果（`gpt-5-mini`）

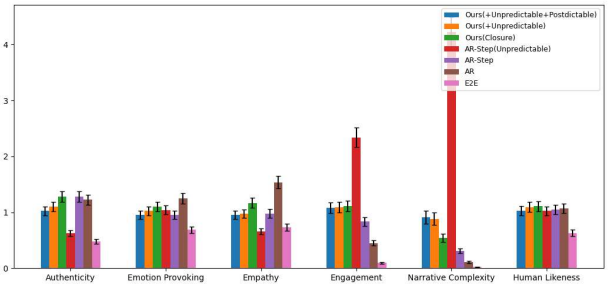

図 5　*PDS* での評価結果（`gpt-5-mini`）

Psychological Depth Scale　以下の 6 つの側面について評価を行う [5]．元々は絶対評価の指標であるが，評価の安定性を高めるためペアワイズ比較用に改変し使用している．

- Authenticity: 人間の体験や心理描写にリアリティや親近感を感じられるかを評価する．
- Emotion Provoking: 強い感情反応を引き起こす能力があるかを評価する．
- Empathy: 登場人物に没入でき，その体験や感情を我がことのように共有できるかを評価する．
- Engagement: 注意を惹きつけ，物語の世界に関心を持たせ続ける力があるかを評価する．
- Narrative Complexity: プロットや人物造形が重層的で，知的な興味や解釈を促すかを評価する．
- Human Likeness

E　事後説明可能性の保証の影響

`gemini-3-flash-preview` を用いた実験において，提示する順序によらず Ours (+Unpredictable+Postdictable) の *Agents' Room Metrics* の Plot スコアが一貫して Ours (+Unpredictable) より低く評価された事例 39 組について，その要因を分析した．

過半数を占める 21 組において，結末で唐突に新たな人や物が登場するなど論理的整合性を欠く解決策が提示されていることがわかった．また，6 組については「一度拒絶した仕事依頼が，再度全く同じ内容で提案される」といった，過去のイベントとの不自然な重複が確認された．以上のことから，過去のミクロ質問に基づいた解決を強制することで，作為的な解決策になったり，過去の文脈を不適切に繰り返したりする悪影響が生じたと推察される．