文脈内知識は LLM の信念体系に整合的に統合されるか？

丹羽 彩奈　金子 正弘　乾 健太郎
MBZUAI
{Ayana.Niwa, Masahiro.Kaneko, Kentaro.Inui}@mbzuai.ac.ae

概要

大規模言語モデル（LLM）は学習を通じて信念同士が相互に結びついた信念体系を獲得する。学習以降の世界の変化に対応するため、新規知識を文脈として与える手法が広く用いられているが、新規知識の獲得だけでなく、それを既存の信念体系へ整合的に統合することが実応用上不可欠である。本研究では、文脈として与えられた新規知識がモデルの信念体系に整合的に統合されるかを評価するベンチマーク TempRIPPLE を提案する。実験結果より、現行の LLM が新規知識を信念体系に整合的に統合できていないことが示唆された。

1　はじめに

大規模言語モデル（LLM）は、大規模コーパスからの事前学習を通じて、世界に関して LLM が真であるとみなす命題、すなわち**信念**を獲得する [1, 2, 3]。これらの信念はモデルのパラメータに符号化され、論理的・因果的・意味的な依存関係を通じて相互に結びつき、ネットワークとしての信念体系をなす [4]。現実世界の事実は刻々と変化するため、LLM は学習以降に生じた変化を信念体系へ動的に反映する能力が求められる [5, 6, 7]。

この課題に対処するために、モデルの学習以降に生じた新規知識[1)]を文脈として与え、動的に信念体系へ反映させる手法が提案されている [8, 9, 10]。しかし、信念がネットワークをなす以上、新規知識の導入は、新規信念の形成にとどまらず周辺の関連信念へ連鎖的に波及しうる [11]。そのため、文脈を通じて新規知識を与える際には、新規信念の獲得だけでなく、既存信念の適切な保持と、関連信念への整合的な波及が求められる。例えば、図 1に示すように、「2025 年のアメリカ大統領は Donald Trump である」という新規知識を与えた場合、モデルは同時に「前大統領は Joe Biden である」という既存信念も保持し、さらに「前大統領の配偶者は Jill Biden である」という関連信念へも適切に波及させる必要がある。本研究では、この過程を新規知識の信念体系への**統合**と呼ぶ。

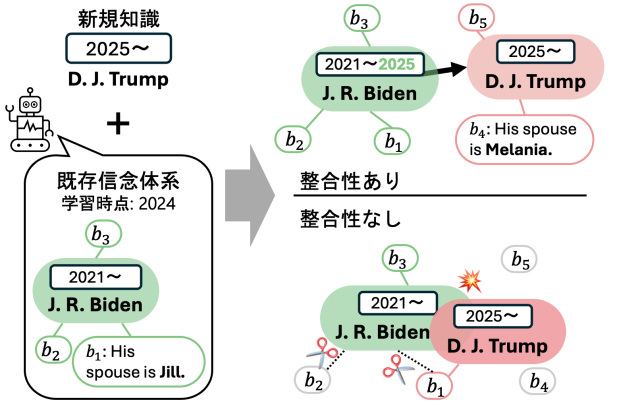

図 1: 新規知識が与えられた際、モデルは新規信念を正しく獲得しつつ、既存信念を保持し、既存関連信念・新規関連信念へも適切に波及させる必要がある。

先行研究は、新規知識そのものの理解（獲得）や、既存知識との矛盾時に新規知識を優先する改訂に焦点を当てており [12, 13]、既存の信念体系への波及は考慮されていない。しかし、時系列的な事実の変化において、既存知識は過去の時点で正当であった知識であり、単純に棄却すべきものではない。新規知識のみを評価対象とした場合、既存の信念体系を破壊しながらも新規知識には正しく応答するモデルを誤って高評価するリスクがある。また整合性が失われた信念体系での推論は幻覚や論理的破綻を招く危険性がある [14]。新規知識を既存の信念体系へ整合的に統合することは実応用において不可欠である。にもかかわらず、既存研究ではこの能力の検証も評価枠組みの整備も十分になされていない。

本研究では、文脈として与えられた新規知識をモデルが既存の信念体系へ整合的に統合できるかを評価する、合計約 37,000 件からなる QA ベンチ

1) **知識**は信念とは異なり、事実として正しい命題を指す。

マーク **TempRIPPLE** を提案する。TempRIPPLE は、Wikidata から抽出した実世界の時間変化に基づく知識を用いたベンチマークである。新規知識を文脈として与えた際に、(1) 新規信念を正しく獲得できるか、(2) 既存信念を保持できるか、(3, 4) 新規あるいは既存それぞれの関連信念へ多段階的に波及できるか、という 4 観点から統合能力を評価する。

4 つの LLM（Qwen2、Qwen2.5、OLMo、GPT-4o-mini）を用いた実験の結果、モデルは新規信念の獲得には高い正解率（95%以上）を示す一方、既存信念の保持では正解率が有意に低下し、関連信念への波及ではホップ数の増加に伴い性能がさらに低下した。これは、現行の LLM が新規知識を獲得できても、それを既存の信念体系に整合的に統合する能力が不十分であることを示唆する。

2 TempRIPPLE ベンチマーク

2.1 タスク定義

知識を $k = (s, p, o)$ と表記する。ここで、$s \in \mathcal{E}$ はサブジェクト、$p \in \mathcal{P}$ はプロパティ、$o \in \mathcal{E} \cup \mathcal{L}$ はオブジェクト、$\mathcal{P}$ はプロパティ集合、$\mathcal{E}$ はエンティティ集合、$\mathcal{L}$ はリテラル集合である。

新規知識と既存知識　モデル M の学習終了時点を t_M としたとき、**新規知識** $k_{\text{new}} = (s, p, o_{\text{new}})$ は時刻 $t > t_M$ において成立する知識であり、**既存知識** $k_{\text{old}} = (s, p, o_{\text{old}})$ は時刻 $t' \le t_M$ において成立していた知識である。両者は同一のサブジェクトとプロパティ (s, p) を共有し、時間変化によってオブジェクトが o_{old} から o_{new} へ遷移する。例えば、学習終了時点が 2024 年のモデルに対して、アメリカ大統領に関する知識は $k_{\text{old}} =$ (アメリカ, 大統領, Joe Biden) および $k_{\text{new}} =$ (アメリカ, 大統領, Donald Trump) となる。

関連知識　n ホップの関連知識 $k^{(n)} = (s^{(n)}, p^{(n)}, o^{(n)})$ は、前ホップのオブジェクトをサブジェクトとする知識 $s^{(n)} = o^{(n-1)}$,　$k^{(0)} = k$ で再帰的に定義される。例えば、先の例における 1 ホップ関連知識は、$k^{(1)}_{\text{old}} =$ (Joe Biden, 出身地, スクラントン) と $k^{(1)}_{\text{new}} =$ (Donald Trump, 出身地, ニューヨーク) である。k が新規知識 k_{new} の場合は新規関連知識、既存知識 k_{old} の場合は既存関連知識となる。

評価タスク　本ベンチマークでは、以上のように構築した知識を問う QA タスクを採用し、新規知識を文脈として与えた際の質問応答を通じて、以下のように各知識の統合能力を評価する。

(1) 新規知識 QA　新規信念を正しく**獲得**できるか
(2) 既存知識 QA　既存信念を正しく**保持**できるか
(3) 新規関連知識 QA　新規信念から関連信念へ正しく**波及**できるか
(4) 既存関連知識 QA　既存信念から関連信念へ正しく**波及**できるか

2.2 構築方法

TempRIPPLE は、2015 年 1 月 1 日から 2025 年 12 月 31 日までの 11 年間を対象として Wikidata [15] をもとに構築した。詳細は付録に記載した。

プロパティの選定　時間変化を伴う新規・既存知識については、先行研究 [16] を参考に、変化頻度が高いと考えられる 16 種類のプロパティを選定した（表 4）。選定したプロパティには、行政府の長（P6）、スポーツチーム所属（P54）、CEO（P169）などが含まれる。一方、関連知識については、誕生日や設立日など不変である可能性が高い 53 種類のプロパティを先行研究 [17] に基づき選定した（表 5）。

可変知識の抽出　2025 年 1 月 1 日から 12 月 31 日までの期間に値が変化した新規知識を Wikidata SPARQL エンドポイントから抽出した。抽出条件として、開始・終了時刻修飾子（`pq:P580`、`pq:P582`）の少なくとも片方が対象期間内にある事実を取得した。次に、抽出した各新規知識について 2015 年から 2024 年までの変化履歴（既存知識）を同じく Wikidata SPARQL エンドポイントから取得した。

関連知識の収集　抽出した各新規・既存知識について、最大 $n = 3$ ホップまで関連知識を探索した。

QA データの作成　抽出した知識をテンプレート（表 4、表 5）を用いて自然言語の QA 形式に変換した。さらに、新規・既存知識には、時間的順序を示す修飾表現を付与した。新規知識には “the most recent”、既存知識には時系列順に “the previous”、“the third most recent” などの序数表現を付与した。[2)]

2.3 データセット統計

464 件のサブジェクト・プロパティの組み合わせから、合計 1,680 件の可変知識および 35,198 件の関連知識を収集した。図 2にデータセットの具体例を示した。各可変知識は、新規知識（2025 年時点）と既存知識（それ以前）の両方について、1 ホップか

2)　新規知識に対して “current” を用いなかったのは、モデルが現在時刻を認識する能力と文脈から新規知識を統合する能力の切り分けを明確にするためである。

	(**開始**, **終了**, **サブジェクト**, **プロパティ**, **オブジェクト**) / **質問** → **正答**
新規知識	(2025, null, United States, head_of_government, **Donald Trump**)
知識 QA	Who is the most recent US president? → **Donald Trump**
関連知識 QA	[Who is the spouse of$_1$ the most recent US president]? → **Melania Trump** [What is the native language of$_2$ [the spouse of$_1$ the most recent US president]]? → **Slovene** [What is the ISO code of$_3$ [the native language of$_2$ [the spouse of$_1$ the most recent US president]]]? → **sl**
既存知識	(2021, 2025, United States, head_of_government, **Joe Biden**)
知識 QA	Who was the previous US president? → **Joe Biden**
関連知識 QA	[Who is the spouse of$_1$ the previous US president]? → **Jill Biden** [Who is the mother of$_2$ [the spouse of$_1$ the previous US president]]? → **Bonny Jacobs** [Where was born$_3$ [the mother of$_2$ [the spouse of$_1$ the previous US president]]]? → **New Jersey**

図 2: TempRIPPLE の例。角括弧内の添字はホップ数を示す（1 ホップ、2 ホップ、3 ホップ）。

表 1: プロパティごとの可変知識の件数と継続期間の統計（抜粋）。事実数は (s, p) の種類数、事例数は既存知識と新規知識の合計インスタンス数を表す。

ID	Property	事実数	事例数	平均	最大
P6	head of gov.	77	192	6.7	31
P35	head of state	11	23	7.5	25
P54	sports team	138	762	2.3	15
P108	employer	66	161	9.3	55
P169	CEO	18	41	6.7	20
P286	head coach	42	205	2.2	16
P488	chairperson	92	248	5.4	23
... （略）					
合計		464	1,680	-	-

ら 3 ホップまでの質問応答ペアから構成される。なお、関連知識の内訳は、1 ホップが 9,338 件、2 ホップが 16,370 件、3 ホップが 9,490 件である。

表 1に、プロパティごとの可変知識の件数および継続期間の統計（抜粋版）を示す。スポーツチーム所属（P54）は平均継続期間 2.3 年で 138 件、ヘッドコーチ（P286）は平均 2.2 年で 42 件と、全プロパティで平均 4.3 年ごとに変化する事実が収集された。

3 実験

3.1 実験設定

TempRIPPLE に含まれる新規知識は 2025 年 1 月以降の情報であるため、学習終了時点が 2024 年中にあるモデルとして、オープンモデル 3 種（Qwen2-7B[3]、Qwen2.5-7B[4]、 OLMo-7B[5] ）とクローズドモデル 1 種（GPT-4o-mini[6]）を採用した。タスクは 4 択の選択式 QA とした。正解以外の選択肢は、同一プロパティを持つ別事例のオブジェクトからランダムに選択した。評価指標には正解率を用いた。チャンスレートは 25%である。新規知識の統合能力を調べるため、各 QA タスクについて、新規事実を自然言語文に変換した新規知識（例 “The most recent US president is Donald Trump.”）を文脈として与えた条件（w/ ctx）と与えない条件（w/o ctx）の 2 条件で評価を行った。前者が統合を要する設定である。

3.2 結果

表 2に、各 QA タスクの正解率を示す。関連知識のホップ数は 1 である。

(1) 新規信念の獲得能力 新規知識 QA において、文脈ありでは全モデルで 95%以上の正解率を達成した。これは、文脈なしと比較して 32〜44 ポイント高い結果となった。本タスクは文脈のみから回答可能であるため高い正解率は想定通りであるが、モデルは新規信念を正しく獲得できることがわかった。一方で、全モデルの正解率が 100%に達していないことから、文脈として与えられた新規知識を常に受け入れるわけではないことも明らかになった。

(2) 既存信念の保持能力 既存知識 QA では、文脈ありで全モデルにおいて正解率が 1〜3 ポイントと、ほぼ全てのモデルで有意に低下した。新規知識は既存知識と時間軸上で矛盾しないにもかかわらず、新規知識を文脈として与えることで既存信念が損なわれることがわかった。

3） Qwen/Qwen2-7B-Instruct
4） Qwen/Qwen2.5-7B-Instruct
5） allenai/OLMo-7B-0724-Instruct-hf
6） gpt-4o-mini-2024-07-18

表 2: 新規知識を文脈として与えた場合（w/ ctx）と与えない場合（w/o ctx）の正解率。†: ホップ間で有意差あり（フィッシャーの正確確率検定、$p < 0.05$）、‡: 同一ホップ内の文脈有無間で有意差あり（マクネマー検定、$p < 0.05$）。

	新規知識 QA		新規関連知識 QA		既存知識 QA		既存関連知識 QA	
モデル	w/o ctx	w/ ctx	w/o ctx	w/ ctx	w/o ctx	w/ ctx	w/o ctx	w/ ctx
Qwen-2	.569	.976‡ ↑.41	.538	.819†‡ ↑.28	.674	.667 ↓.01	.717†	.636†‡ ↓.08
Qwen-2.5	.594	.982‡ ↑.39	.520	.835†‡ ↑.31	.716	.695‡ ↓.02	.642†	.618†‡ ↓.02
OLMo	.557	.994‡ ↑.44	.508†	.795†‡ ↑.28	.596	.580‡ ↓.02	.552†	.530†‡ ↓.02
GPT-4o	.634	.958‡ ↑.32	.585†	.819†‡ ↑.23	.719	.688‡ ↓.03	.631†	.601†‡ ↓.03

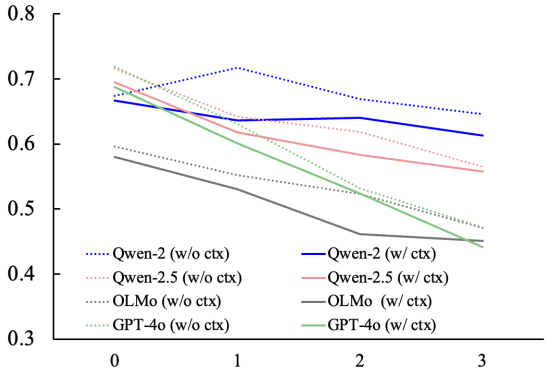

図 3: 既存知識（0 と表記）から既存関連知識（ホップ数の 1, 2, 3 で表記）の QA における正解率。

(3,4) 関連信念への波及能力　関連知識 QA では、新規・既存知識それぞれと比較して正解率が有意に低下した。文脈ありで新規知識 QA では最大 20 ポイント、既存知識 QA では最大 8 ポイントの低下が見られた。つまり、新規信念を獲得できたとしても、そこからの関連信念への波及はできていない。

これらの 4 つの結果から、現行の LLM は信念体系への統合が十分にできていないことを示唆する。

4　分析

信念の統合は体系全体に波及するか？　図 3に既存関連知識（1, 2, 3 ホップ）の正解率の推移を示したように、ホップ数の増加に伴い、多くのケースで正解率が低下し、既存知識に比べて 3 ホップでは最大 25 ポイント低下した。これが単なる多段階的な推論の難しさによるものであれば、文脈の有無に依存しない一様な性能低下が予想される。しかし、関連知識においては全てのモデルで文脈の有無に有意な差が見られた（マクネマー検定、$p < 0.05$）。つまり、新規知識が信念体系全体へ整合的に波及しているのではなく、局所的な知識更新に留まっている。

知識の定着度は統合のしやすさに影響するか？　認識論によると、人間がもつ信念にはそれぞれ定着度があり、論理や常識といった変化しにくい知識ほど信念も変更されにくい [4, 18, 19]。そこで LLM がこのような定着度に基づく変更の優先順位を持ち合わせているか検証するため、プロパティの平均継続年数と正解率の相関（点双列相関係数）を分析した。結果、表 3に示したように、新規・既存知識 QA ともに相関係数は 0.12 以下と低く、ほとんど相関が見られなかった。これは、事実の変化頻度に応じた定着度の構造が少なくとも本設定では観測されず、人間のような合理的な変更の優先順位付けが行われているとは言い難いことを示唆している。これは、定着度の高いドメインとして科学事実を採用し、同能力を検証した先行研究 [19] と整合的な知見である。

表 3: プロパティごとの平均継続年数と文脈あり QA における正解率の相関係数。

モデル	新規知識	既存知識
Qwen-2	0.0159	0.1169
Qwen-2.5	0.0652	0.1175
OLMo	0.0560	0.1192
GPT-4o	0.0077	-0.0001

5　おわりに

本研究では、LLM が文脈として与えられた新規知識を既存の信念体系へ整合的に統合できるかを評価するベンチマーク TempRIPPLE を提案した。TempRIPPLE は、Wikidata から抽出した実世界の時間変化に基づく知識を用い、新規信念の獲得、既存信念の保持、関連信念への波及の観点から統合挙動を評価する。実験の結果、LLM は新規信念の獲得には高い正解率を示す一方で、既存信念の保持や関連信念への波及では低い正解率を示した。これらの結果は、獲得した新信念が既存の信念体系へ整合的に統合する能力が不十分であることを示唆している。

参考文献

[1] Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. Language models as knowledge bases? In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, **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. 2463–2473, Hong Kong, China, November 2019. Association for Computational Linguistics.

[2] Bryan Wilie, Samuel Cahyawijaya, Etsuko Ishii, Junxian He, and Pascale Fung. Belief revision: The adaptability of large language models reasoning. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing**, pp. 10480–10496, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[3] Nora Kassner, Oyvind Tafjord, Hinrich Schütze, and Peter Clark. BeliefBank: Adding memory to a pre-trained language model for a systematic notion of belief. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 8849–8861, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.

[4] Willard Van Orman Quine. Two dogmas of empiricism. **Perspectives in the Philosophy of Language**, pp. 189–210, 2000.

[5] Angeliki Lazaridou, Adhi Kuncoro, Elena Gribovskaya, Devang Agrawal, Adam Liska, Tayfun Terzi, Mai Gimenez, Cyprien de Masson d'Autume, Tomas Kocisky, Sebastian Ruder, et al. Mind the gap: Assessing temporal generalization in neural language models. **Advances in Neural Information Processing Systems**, Vol. 34, pp. 29348–29363, 2021.

[6] Bhuwan Dhingra, Jeremy R Cole, Julian Martin Eisenschlos, Daniel Gillick, Jacob Eisenstein, and William W Cohen. Time-aware language models as temporal knowledge bases. **Transactions of the Association for Computational Linguistics**, Vol. 10, pp. 257–273, 2022.

[7] Jungo Kasai, Keisuke Sakaguchi, Ronan Le Bras, Akari Asai, Xinyan Yu, Dragomir Radev, Noah A Smith, Yejin Choi, Kentaro Inui, et al. Realtime qa: What's the answer right now? **Advances in neural information processing systems**, Vol. 36, pp. 49025–49043, 2023.

[8] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. **Advances in neural information processing systems**, Vol. 33, pp. 9459–9474, 2020.

[9] Gautier Izacard and Edouard Grave. Leveraging passage retrieval with generative models for open domain question answering. In **Proceedings of the 16th conference of the european chapter of the association for computational linguistics: main volume**, pp. 874–880, 2021.

[10] Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christopher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, et al. Webgpt: Browser-assisted question-answering with human feedback. **arXiv preprint arXiv:2112.09332**, 2021.

[11] Song Wang, Yaochen Zhu, Haochen Liu, Zaiyi Zheng, Chen Chen, and Jundong Li. Knowledge editing for large language models: A survey, 2024.

[12] Xiaowei Yuan, Zhao Yang, Yequan Wang, Shengping Liu, Jun Zhao, and Kang Liu. Discerning and resolving knowledge conflicts through adaptive decoding with contextual information-entropy constraint. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, **Findings of the Association for Computational Linguistics: ACL 2024**, pp. 3903–3922, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

[13] Qinggang Zhang, Zhishang Xiang, Yilin Xiao, Le Wang, Junhui Li, Xinrun Wang, and Jinsong Su. FaithfulRAG: Fact-level conflict modeling for context-faithful retrieval-augmented generation. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, **Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 21863–21882, Vienna, Austria, July 2025. Association for Computational Linguistics.

[14] Rongwu Xu, Zehan Qi, Zhijiang Guo, Cunxiang Wang, Hongru Wang, Yue Zhang, and Wei Xu. Knowledge conflicts for LLMs: A survey. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing**, pp. 8541–8565, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[15] Denny Vrandečić and Markus Krötzsch. Wikidata: A free collaborative knowledgebase. **Commun. ACM**, Vol. 57, No. 10, pp. 78–85, September 2014.

[16] Keyuan Cheng, Gang Lin, Haoyang Fei, Yuxuan Zhai, Lu Yu, Muhammad Asif Ali, Lijie Hu, and Di Wang. Multi-hop question answering under temporal knowledge editing. In **First Conference on Language Modeling**, 2024.

[17] Roi Cohen, Eden Biran, Ori Yoran, Amir Globerson, and Mor Geva. Evaluating the ripple effects of knowledge editing in language models. **Transactions of the Association for Computational Linguistics**, Vol. 12, pp. 283–298, 2024.

[18] Willard Van Orman Quine and J. S. Ullian. **The Web of Belief**. Random House, New York,, 1970.

[19] Minsu Kim and James Thorne. Epistemology of language models: Do language models have holistic knowledge? In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, **Findings of the Association for Computational Linguistics: ACL 2024**, pp. 12644–12669, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

表 4: 可変事実の作成に用いたテンプレート。

プロパティ	命題テンプレート	質問テンプレート
P35	The head of state in <subject> is <object>	Who is the head of state in <subject>?
P6	The head of government in <subject> is <object>	Who is the head of government in <subject>?
P488	The chairperson of <subject> is <object>	Who is the chairperson of <subject>?
P169	The chief executive officer of <subject> is <object>	Who is the chief executive officer of <subject>?
P54	<subject> is affiliated with the sports team of <object>	Which sports team is <subject> affiliated with?
P286	The head coach of <subject> is <object>	Who is the head coach of <subject>?
P551	<subject> lives in <object>	Where does <subject> live?
P102	<subject> is affiliated with the political party of <object>	Which political party is <subject> affiliated with?
P108	<subject> is an employee of <object>	Which organization is <subject> an employee of?
P69	<subject> is educated at <object>	Which university is <subject> educated at?
P937	The work location of <subject> is <object>	Where is <subject>'s workplace?
P36	The capital of <subject> is <object>	What is the capital of <subject>?
P159	The headquarters of <subject> is located in <object>	Where is the headquarters of <subject> located?
P27	<subject> is a citizen of <object>	What is the country of citizenship of <subject>?
P140	<subject> is affiliated with the religion of <object>	Which religion is <subject> affiliated with?
P17	<subject> is located in the country of <object>	Which country is <subject> located in?

表 5: 関連事実の作成に用いたテンプレート（抜粋版）。

プロパティ	テンプレート
P25	Who is the mother of <subject>?
P22	Who is the father of <subject>?
P19	What is the birthplace of <subject>?
P103	What is the native language of <subject>?
P37	What is the official language of <subject>?
P36	What is the capital of <subject>?
P37	What is the official language of <subject>?
P112	Who is the founder of <subject>?
P571	What is the founding date of <subject>?
P452	What is the industry of <subject>?
P115	What is the home venue of <subject>?
P498	What is the ISO code of <subject>?
P361	What is the language family of <subject>?
P282	What is the writing system of <subject>?

A　データ作成の詳細

可変事実の収集時は、プロパティごとの抽出上限を 100 万件とした。

また、対象期間全体を通じて同一期間での重複がなく、単一オブジェクトのみを持ち、1 年以上のデータの欠損のない可変事実のみを採用した。

関連知識の収集時には、不変プロパティをもつことに加え、また実際に 2015 年 1 月 1 日以前に開始し 2025 年 12 月 31 日以降まで継続している、すなわち対象期間を通じて不変である事実のみを対象とした。マルチホップのデータを作成する際は、テンプレートベースで質問を生成するため、各ホップでは起点エンティティの型に対応するプロパティのみを辿った。また、ホップは以下の条件に該当するものを除外した。

(1) 推論パスが循環するもの（起点のサブジェクトまたはオブジェクトが再度出現する場合）

(2) 推論パス内で同一エンティティが複数回出現するもの（連続出現を含む）。