

# J-PIQA: Japanese Physical Interaction Question Answering

Anonymous ACL submission

## 1 Methods

Our dataset consists of two parts, both designed to adapt the PIQA (Physical Interaction Question Answering; Bisk et al., 2020) framework to Japanese language and cultural contexts.

### 1.1 Translation and Filtering of PIQA Subset

We began by randomly sampling 200 items from the original PIQA dataset. To ensure minimal lexical variation between alternatives, we automatically filtered the items to retain only those where the two candidate solutions differed in at most two words, as determined by word-level edit distance. We further excluded examples judged to reflect personal preferences rather than physical reasoning. The resulting subset was translated into Japanese using ChatGPT, and each translation was subsequently reviewed and corrected by two native speakers to guarantee naturalness and fidelity to the source. The final translation set amounts to 145 examples.

### 1.2 Generation of Culture-Specific Variants

To introduce culturally grounded content, we explored two strategies. First, we modified the translated PIQA items by prompting ChatGPT to replace lexical elements with counterparts specific to Japanese culture (e.g., food, customs, and everyday objects). Of the 145 examples from the translation set, 35 were deemed appropriate by the 2 native speakers of Japanese. For broader coverage, we then prompted ChatGPT to generate 300 new PIQA-style examples directly in Japanese, constrained such that the two candidate solutions differed in at most two words and required knowledge of Japanese cultural norms and conventions for resolution. From this pool, 66 items were manually selected to ensure quality and adherence to the intended reasoning task.

Together, this dataset with 246 items (145 translations, 35 lexically-altered items, and 66 original

items) constitute a dataset that both preserves the structure of PIQA and extends it to test physical reasoning in Japanese cultural contexts.

## References

Yonatan Bisk, Rowan Zellers, Jianfeng Gao, Yejin Choi, and 1 others. 2020. Piqa: Reasoning about physical commonsense in natural language. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 7432–7439.

## A Columns

Besides the required columns prompt, solution0, solution1, and label, our data contains the following 6 columns: prompt\_gloss, solution0\_gloss, solution1\_gloss, source, orig\_line\_no, and is\_translation. The first three columns, prompt\_gloss, solution0\_gloss, solution1\_gloss, are the English gloss of the original prompt, solution0, and solution1 columns. label source takes three possible values: piqa\_translation, piqa\_lexical\_variation, and piqa\_original. piqa\_translation is a translated sample of PIQA item (Section 1.1), and the other two are culture-specific variants described in Section 1.2.

orig\_line\_no specifies the original item ID from PIQA if the item was derived from the original PIQA. This column applies when source column is piqa\_translation or piqa\_lexical\_variation. Otherwise, this column will have a value of -1. is\_translation is 1 for translated items, and 0 for others.

## B Prompts

Figure 1 is the Japanese prompt used to generate culture-specific variants, and its translation is shown in Figure 2.

あなたは熟練のデータアノテーターです。日本文化を前提にした物理的推論を測る二択問題を300問作成してください。

目的

- 物理法則、器具の機能、危険回避、日本固有の生活様式や道具や慣習の知識を使わないと正解できない問題にする。
- 好ましさやマナーの優劣ではなく、物理的にどちらが正しいかが一意に決まる問題にする。

出力形式 (CSV/Excel 互換)

ヘッダー (固定):

```
id,category,question,option1,option2,correct,reasoning
[,english_question,english_option1,english_option2,english_reasoning]
```

列の定義

- id: 1..300 の連番
- category: 食べ物, 住居, 生活, 工具, 交通, 防災, 行事, 文化, 音楽, 教育, 趣味/アウトドア など
- question: 日本語の設問。単語のみでも完全文でもよい。短文と長文を混在させる。
- option1, option2: どちらも日本語の完全文で末尾は「。」。両者は 1 から 2 語だけが異なる最小差分。語順や句読点や助詞は同一。
- correct: option1 が正解なら 1, option2 が正解なら 2。一意に決まること。
- reasoning: 正解の物理的理由を日本語で 1 文。日本の文脈に根ざした説明。
- english\_\*: question, option1, option2, reasoning の英訳列

厳守ルール

- 1) 最小差分
- option1 と option2 は 1 から 2 語のみ差し替え。語順や句読点や助詞や体裁は同一。
- 例: ラップをかけて加熱する。 / アルミ箔をかけて加熱する。
- 2) オプションは完全文
- 両オプションとも末尾「。」の完全文。断片は禁止。
- 3) 物理的に正誤が決まる
- 物理法則、材質、道具の適合、危険回避により正解が一意に決まる。マナーや好みでは決めない。
- 4) 日本固有の知識が効く
- 玄関の段差、畳、上履き、網戸、障子、襖、七輪、石油ストーブ、神輿、線香、浴衣、茶道具、和楽器、資源回収などを積極的に扱う。
- 5) 分野バランス
- 食べ物、工具、住居、生活、防災、交通、行事、文化、音楽、教育、趣味/アウトドアをまんべんなく。短文と状況文を混在。
- 6) 重複厳禁
- (question, option1, option2) の完全一致重複をゼロにする。言い換え重複も避ける。
- 7) 危険行為の扱い
- 火気、感電、転落、窒息、一酸化炭素、交差汚染などは回避が正解。油火災に水をかけない。レンジに金属を入れない。ガス臭は換気する。
- 8) 曖昧禁止
- 季節や条件で正解が変わる場合は設問に条件を明記する。例: 真夏の保存はどうするか。
- 9) 言語
- 日本語列は自然な日本語。英訳列は自然な英語。

推奨トピック例

- 食べ物: 真夏の保存、加熱が必要な食材、交差汚染、電子レンジ可否、鍋蓋の開け方、弁当冷却
- 住居: 玄関の靴、畳、障子、襖、除湿機と窓、電源タップのほこり、延長コードの束ね使用禁止
- 防災と火気: 七輪は屋外、石油ストーブは換気、台風前の固定、地震時の避難路、油鍋着火の対応
- 交通: 踏切、ホームの黄色線、ベビーカーはエレベーター、夜間ライト、雪道タイヤやチェーン
- 文化と行事: 神社参拝、手水、箸を立てない、神輿の足元、線香花火の向き、浴衣の合わせ
- 教育と趣味: 体育館は上履き、理科の保護具、登山の熊鈴、釣りのライフジャケット
- 工具: 材質に適合する工具、包丁の受け渡し、刃の鞘、砥石の水、火花の方向

品質チェック

- 300 行あるか。id は 1..300 の連番か。
- 正解が各行で一意か。correct は 1 または 2 か。
- option1 と option2 は 1 から 2 語差のみで完全文か。末尾は「。」か。
- 理由は日本語 1 文で物理的根拠が明確か。
- 重複なし。
- 分野バランスが良いか。
- 曖昧や好みやマナー優先になっていないか。

最終出力要件

- 上記ヘッダーの CSV 表のみを出力する。この指示テキストは出力しない。

Figure 1: Original Japanese prompt supplied to ChatGPT for the generation of culture-specific variants.

You are an expert data annotator. Create 300 two choice questions that test physical reasoning grounded in Japanese culture.

Purpose

- Each question must require knowledge of physics, mechanics, safety, and Japanese specific living customs, tools, and practices.
- This is not about preferences or etiquette. The correct answer must be physically correct and uniquely determined.

Output format (CSV or Excel compatible)

Header (fixed):

```
id,category,question,option1,option2,correct,reasoning
[,english_question,english_option1,english_option2,english_reasoning]
```

Column definitions

- id: integers 1..300
- category: Food, Housing, Daily life, Tools, Transport, Disaster safety, Events, Culture, Music, Education, Hobby or Outdoor, etc.
- question: The prompt in Japanese. It may be a single word or a full sentence. Mix short items and longer scenarios.
- option1, option2: Complete Japanese sentences ending with the full stop symbol. They must differ by only 1 to 2 words. Keep order and punctuation and particles identical.
- correct: 1 if option1 is correct, 2 if option2 is correct. Must be unique.
- reasoning: One sentence physical rationale in Japanese tied to the Japanese context.
- english\_\*: English translations for question, option1, option2, reasoning.

Hard constraints

- 1) Minimal difference
- Options differ by only 1 to 2 words. Keep formatting and order identical.
- Example: Wrap with plastic film and heat. / Wrap with aluminum foil and heat.
- 2) Complete sentences
- Both options are complete Japanese sentences ending with a full stop symbol.
- 3) Physically determinate
- Correctness follows from physics, material suitability, and safety, not etiquette or taste.
- 4) Japan specific knowledge
- Use topics that require Japanese cultural knowledge such as genkan step, tatami, indoor shoes, shoji, fusuma, charcoal stove, kerosene heater, mikoshi, incense, yukata, tea tools, traditional instruments, recycling rules.
- 5) Broad topical coverage
- Balanced mix across Food, Tools, Housing, Daily life, Disaster, Transport, Events, Culture, Music, Education, Hobby or Outdoor. Mix short and scenario style prompts.
- 6) No duplicates
- Zero duplicate rows of the triplet question, option1, option2. Avoid paraphrase duplicates as well.
- 7) Hazard handling
- Safety preserving behavior should be correct. Do not pour water on oil fires. Do not put metal in a microwave. Ventilate when you smell gas.
- 8) No ambiguity
- If correctness depends on season or appliance, state it in the question. Example: In midsummer, how to store the item.
- 9) Language
- Natural Japanese for JP fields and natural English for the english columns.

Suggested topics

- Food: midsummer storage, foods requiring thorough cooking, cross contamination, microwave safety, opening pot lids, cooling bento
- Housing: genkan shoes, tatami, shoji and fusuma, dehumidifier versus windows, dust on power strips, no coiled extension cords
- Fire and disaster: outdoor only charcoal devices, ventilation with kerosene heaters, securing balconies before typhoons, keeping egress open after earthquakes, smothering oil fires
- Transport: rail crossings and yellow safety lines, elevators for strollers, night lights, snow tires or chains
- Culture and events: shrine visits, hand cleansing, do not leave chopsticks upright, footwear for mikoshi, sparkler orientation, yukata left over right
- Education and hobbies: indoor shoes in gyms, lab protective equipment, bear bells in mountains, lifejackets for fishing
- Tools: material appropriate tools, knife handover, sheaths, wet whetstones, spark direction

Quality checklist

- Exactly 300 rows with ids 1..300
- Unique correct value per row (1 or 2)
- Options differ by 1 to 2 words and are full sentences ending with a full stop symbol
- Reasoning is one sentence in Japanese with a clear physical basis
- No duplicate triplets of question, option1, option2
- Balanced topical coverage
- No ambiguity or preference or etiquette based items

Final output requirement

- Output only the CSV table with the header above. Do not print this instruction text.

Figure 2: Translation of the Japanese prompt supplied to ChatGPT for the generation of culture-specific variants.