PiQA- Hebrew

Leshem Choshen

July 2025

1 Introduction

This paper reports on the methodological decisions underlying the curation of a Hebrew adaptation of PiQA [1]. The dataset consists of 192 example questions formulated in Hebrew, each addressing a concrete and well-defined aspect of the physical world. For every question, two candidate answers are provided, of which one is consistently more correct, and ideally unambiguously correct, within the context of Hebrew linguistic and cultural conventions. As the questions are grounded in assumptions of Hebrew language competence and Israeli cultural knowledge, certain items may resist direct translation, and in some cases, translation may alter the validity of the designated correct answer.

Data creation

The whole dataset was curated manually. Overlap between instances was kept to a minimum, except between the answers where as much as possible was shared between the answers and ideally, the change is in towards the end of the answer. The format was kept to sentences which are the outcome and answers which are the cause (rather than a question and other loose formats common in PIQA).

The dataset is curated based on notions that served as seeds for topics. Those topics vary from regional interests, general physical notions or linguistic peculiarities.

Many of the examples ask about culturally specific events. Specifically, 4 examples discuss roads, 6 discuss Places in Israel, 34 discuss food, 13 discuss the Israeli climate, and 11 discuss Jewish religion and culture.

A large portion of the examples dealt with physical questions that focused on aspects of reality rather than culture. This included 14 questions on sizes of things, 10 about textures, 8 about colors, 12 about angles, 16 about magnetism, 10 about sounds, and 12 about special physical effects that are often confusing or unintuitive until learnt or found by modern physics(e.g. quantum frequency), additional 8 examples that are similar to the special physics phenomena but describe a mundane scenario where more common causes explain the behavior are included to prevent contamination of explaining everything with the rare solutions. This is a possibility because those are more salient and

explicitly discussed than common explanations. The special physics examples were created with an LLM and edited by a human.

About 20 questions are about cutting and breaking various physical objects, those are known to be a difficulty in linguistic verb representation and to vary across languages [3, 2], thus, they might be of interest for a linguistically specific query for physical world understanding, which is general.

References

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

[2] Asifa Majid, Melissa Bowerman, Miriam van Staden, and James S Boster. The semantic categories of cutting and breaking events: A crosslinguistic perspective. 2007.

[3] Clifton Pye, Diane Frome Loeb, and Yin-Yin Pao. The acquisition of breaking and cutting. In *The proceedings of the twenty-seventh annual child language research forum*, pages 227–236. Center for the Study of Language and Information Stanford, 1996.