MRL 2025 Shared Task on
Multilingual Physical Reasoning Datasets

Maximos Skandalis

September 15, 2025

1 Greek Dataset Creation

The categories covered by the Greek dataset on physical commonsense reasoning include: vocabulary-relevant puzzles/riddles (26 examples), household (16), cooking/recipes (14), driving (12), gardening (9), DIY (7), sports (6), construction/building (3), vacation situations (3), spatiotemporal orientation (2), and dance (2).

The topic categories were selected after reviewing the examples of the development set of the PIQA dataset, and by reflecting on and evaluating the relevance, diversity, and potential of candidate topics for our purposes.

Most of the examples for Greek were created by navigating Greek websites on the internet and searching for sentences on the chosen topics, then adapting them for the task, by modifying one part of the sentences in order to create the second solution, in a way for the task to be challenging. Changes in the second solution can be either short lexical substitutions (1–3 words) or alterations in the order of the words inside the sentence. Some of the short examples were created from scratch, by the author.

This process yielded 100 examples across 11 categories.

2 French Dataset Composition

For French, the longest examples from the PIQA development subset were translated from English to French. These translated sentences are intended to be added alongside examples produced by other MRL contributors for French, thereby expanding the coverage of French within the Global PIQA dataset.