Dataset Description: Spanish. Álvaro Arroyo

Anonymous EMNLP submission

1 Introduction

We present a PIQA-style commonsense dataset written in Spanish as spoken in Spain (*español peninsular* [Peninsular Spanish]), with a deliberate regional cadence toward Madrid and the interior of Castilla y León. The variety targeted can be described as a central/northern peninsular register: it preserves peninsular morphology and usage, including *vosotros* [2nd person plural informal], and lexis such as *ordenador* [computer], *coche* [car], *zumo* [juice], *persiana* [roller shutter/blind]. Prompts embed culturally grounded references, e.g., Metro de Madrid, *botijo* [porous clay water jug], *cocido* [stew] and *lechazo* [roast suckling lamb], meseta (central plateau) climate, and domestic practices. This register is intentionally culturally specific and differs subtly from more neutral or pan-Hispanic usage through vocabulary choices, idiomaticity, and background assumptions.

The dataset contains exactly 100 examples in the PIQA format `[prompt, solution0, solution1, label]`, manually authored and natively checked. As requested, items fall into three overlapping buckets: (i) **physical** reasoning (e.g., trajectories and bounciness, friction, thermal conduction/insulation, greenhouse effect, pressure, lever/torque); (ii) **culturally specific** scenes (traditional foods, utensils, customs, and common urban settings); and (iii) **everyday commonsense** (day-to-day household and safety scenarios). All items are answerable by an average native speaker. To avoid triviality, the two candidate solutions are intentionally close in form: in the vast majority they differ by only one or two words (minimal pairs), with the incorrect alternative plausible yet physically or pragmatically wrong. Length is controlled: *most (prompt + solutions) exceed 25 words*, and some prompts are multi-sentence. Prompts do not all start the same way and typically end with either a question, an incomplete sentence, or a full stop followed by options. Labels are balanced (≈50/50) to avoid class imbalance, and the TSV was shuffled prior to submission. We adhere to the required columns (`prompt`, `solution0`, `solution1`, `label`, with exactly one correct solution).