

Ёлки Палки: A Russian Dataset of Physical Commonsense Reasoning Tasks

Anonymous ACL submission

1 Introduction

We introduce the Ёлки Палки (Yolki Palki) dataset, containing 100 instances of physical commonsense reasoning tasks.

2 Annotator Background

Two native speakers developed the dataset together. One native speaker (Native Speaker #1) was born in Russia in Ростов-на-Дону (Rostov-on-Don) and lived in Russia for 30 years prior to moving to the United States. The second native speaker (Native Speaker #2) was born in the United States and grew up speaking Russian.

3 Methodology

Both native speakers were involved in developing the dataset and collaboratively discussed topics for the entries in the dataset. Prior to beginning data set construction, Native Speaker #2 went through the slides provided on the task and explained the task to Native Speaker #1. Both native speakers carefully studied the slide presenting example instances in English.

For the first 20 instances of the dataset, the native speakers collaboratively discussed topics related to physical reasoning, culturally relevant topics, and ensured that the answers would be obvious to a native speaker of Russian. The following 60 instances in the dataset were curated by Native Speaker #1 and double-checked by Native Speaker #2. The last 20 instances of the dataset were collaboratively curated between the two native speakers, with Native Speaker #2 providing topics and both native speakers discussing the answers.

4 Notes

For words containing the letter ё, we defaulted to e as is done in many types of writing in Russian since native speakers typically can differentiate whether e refers to ё or e based on the context (Timberlake, 2004).

For the culturally relevant instances of the dataset, several of these instances referred to culturally relevant holidays (such as Новый Год (New Year) or foods.

References

Alan Timberlake. 2004. A reference grammar of Russian. Cambridge University Press.