

## Uzbek MRL Dataset

By Sharifa Djurabaeva

### Dataset Collection Methodology

As a native speaker of Uzbek, I manually constructed this dataset, which comprises 101 examples primarily written in standard Uzbek, with some instances of colloquial Uzbek included as well. Each sentence draws inspiration from real-life experiences and commonly used expressions in the language, ensuring their naturalness and relevance. The sentences vary in length from 5 to 100 words, capturing a broad range of linguistic structures and registers.

For each example, I provided two solutions: a correct one that represents the expected grammatical and semantic form, and an incorrect one that illustrates a plausible mistake that native speakers or learners might make. Initially, it was challenging to identify realistic incorrect forms, but over time, I began to recognize and incorporate recurring patterns in common errors—such as misused alternative words, opposing actions, and mistakes in word order.

The table below demonstrates two examples that are similar to what I have collected.

		Correct	Incorrect		Translations of the prompt	Correct solution translations	Incorrect solution translations
Number	prompt	solution0	solution1	label	prompt	solution0	solution1
1	Tandirda non yopayotganingizda, noningiz oqib ketmasin desangiz,	xamiringizda deyarli suv qolmasligi kerak. Xamir mushtlaganingizda, <u>xamiringiz</u> o'rtacha qattiqlikda, va <u>qo'ingizga yopishmaydigan</u> holatda bo'lishi kerak. Shundagina, issiq tandirga nonni yopishtirganingizdan keyin, noningiz oqib tushmaydi.	xamiringizda deyarli suv qolmasligi kerak. Xamir mushtlaganingizda, <u>xamiringiz</u> o'rtacha qattiqlikda, va <u>qo'ingizga qorishmaydigan</u> holatda bo'lishi kerak. Shundagina, issiq tandirga nonni yopishtirganingizdan keyin,	0		xamiringiz qo'lingizga yopishmaydigan - the dough should not be sticking to your hand	xamiringiz qo'lingizga qorishmaydigan - the dough should not be mixing with your hand

			noningiz oqib tushmaydi.				
2	Barmog'ingizga uzugingiz tiqilib qolsa, va uni chiqarolmasangiz, quydagi amalni bagarib ko'ring.	Birinchi sovun oling va uzuk taqilgan barmog'ingizni yaxshilab iliq suvda sovunlang. Biroz kutib, iliq sovunli suvlar uzuk va barmoqlaringiz orasiga kirgandan keyin, uzukni sekin astalik bilan goh chapga goh o'ngga burang. Agar uzuk chiqavermasa, unda yaxshilab uzuk atrofiga sovun suring va uzukni osonlikcha barmog'ingizdan <u>chiqarib</u> oling.	Birinchi sovun oling va uzuk taqilgan barmog'ingizni yaxshilab iliq suvda sovunlang. Biroz kutib, iliq sovunli suvlar uzuk va barmoqlaringiz orasiga kirgandan keyin, uzukni sekin astalik bilan goh chapga goh o'ngga burang. Agar uzuk chiqavermasa, unda yaxshilab uzuk atrofiga sovun suring va uzukni osonlikcha barmog'ingizdan <u>so'rib</u> oling.	0		chiqarib - take out	so'rib - suck out

The dataset is stored in TSV (Tab-Separated Values) format with columns: number, prompt, correct solution incorrect solution, lable and translation of the prompt, translation of the correct solution, incorrect solution, but only for six examples. Other translations included only the words chosen in a correct solution and its replacement in incorrect solution.