# Generating Continuations in Multilingual Idiomatic Contexts
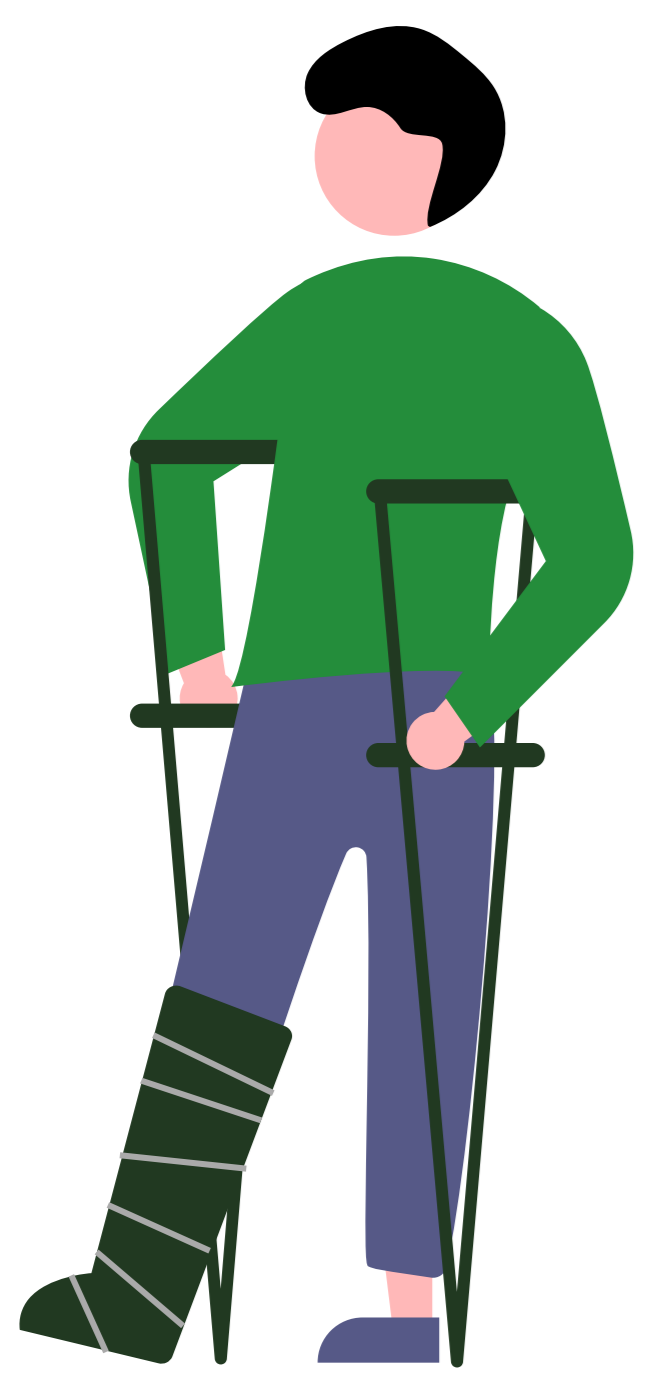
*Rhitabrat Pokharel and Ameeta Agrawal*

*PortNLP Lab, Department of Computer Science, Portland State University*
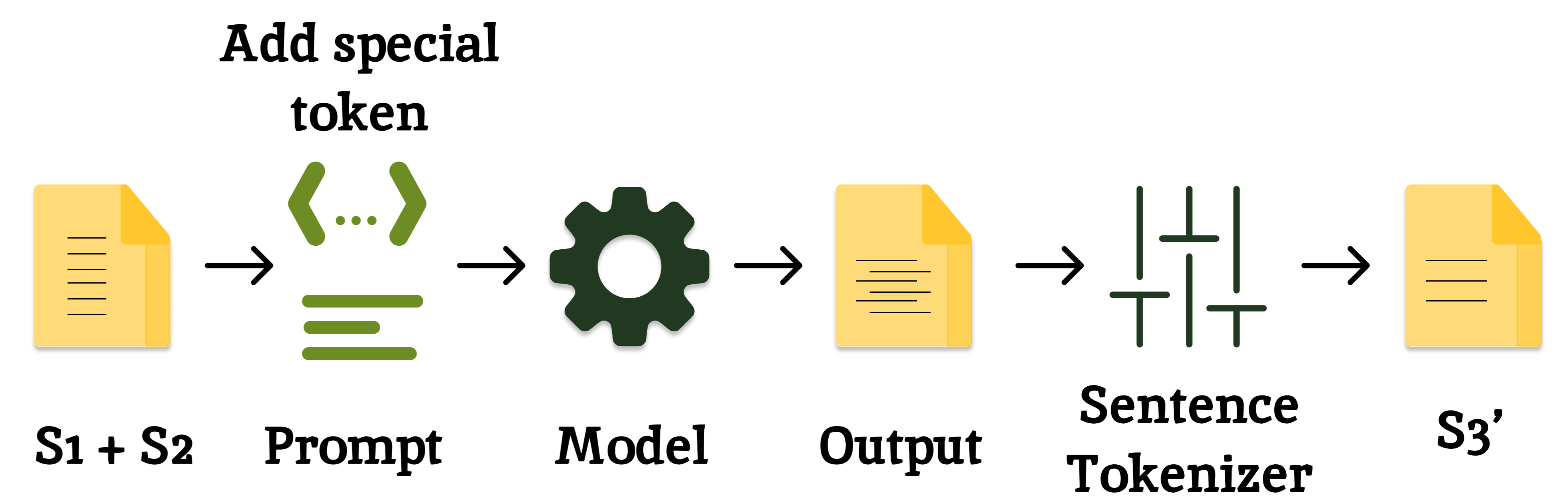
**MRL, EMNLP 2023, Singapore**

## MOTIVATION

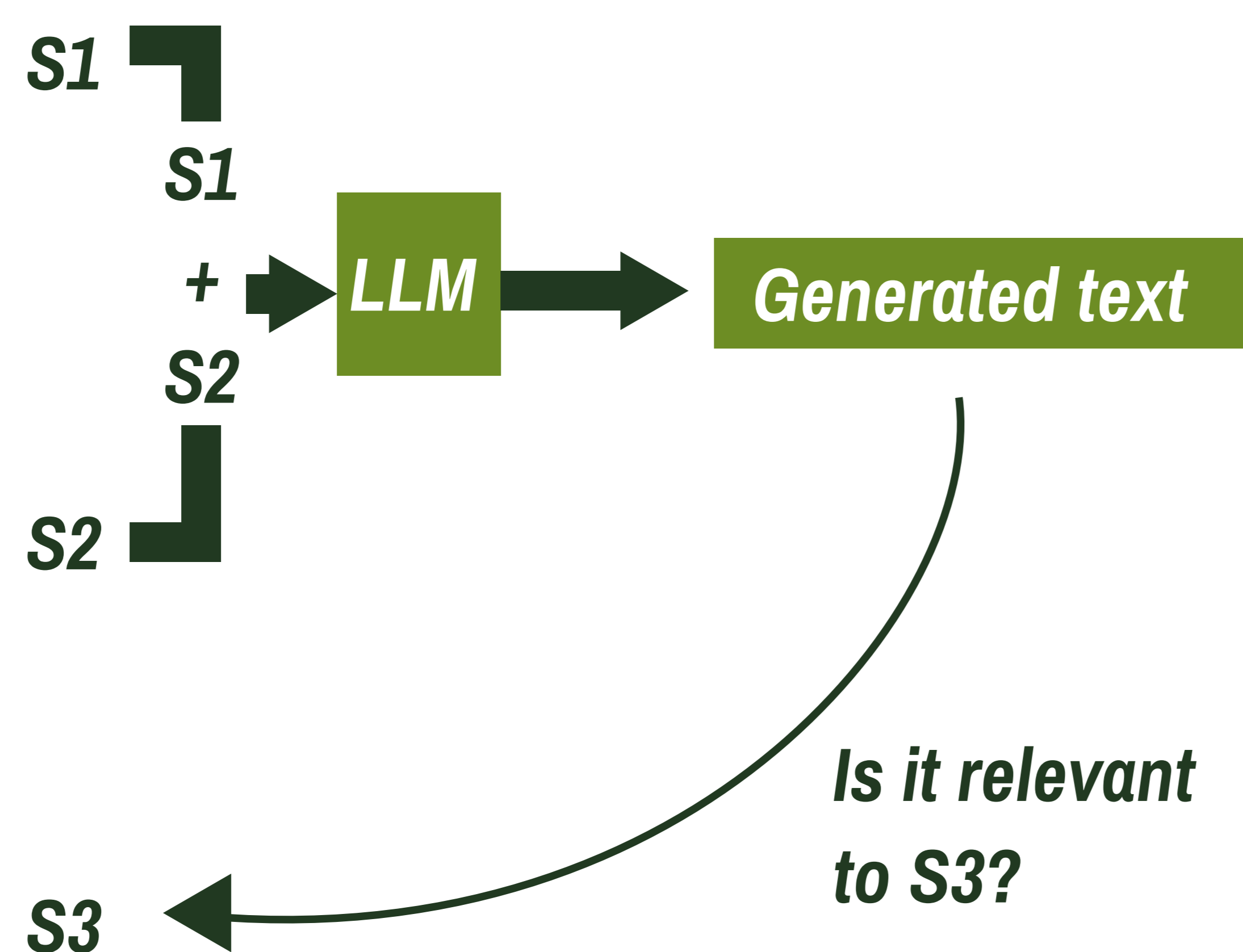*I hope you break a leg in your theater performance tonight!*

## PROBLEM FORMULATION

Gold remains stuck in consolidation mode, and this can be frustrating for some investors.

At times like this, it is critical to remain focused on the big picture.

Our primary forecast still anticipates a minimum target of $8,500 by 2028.

S1

S1 + S2 → LLM → Generated text

S2

S3

Is it relevant to S3?

## DATA SAMPLES

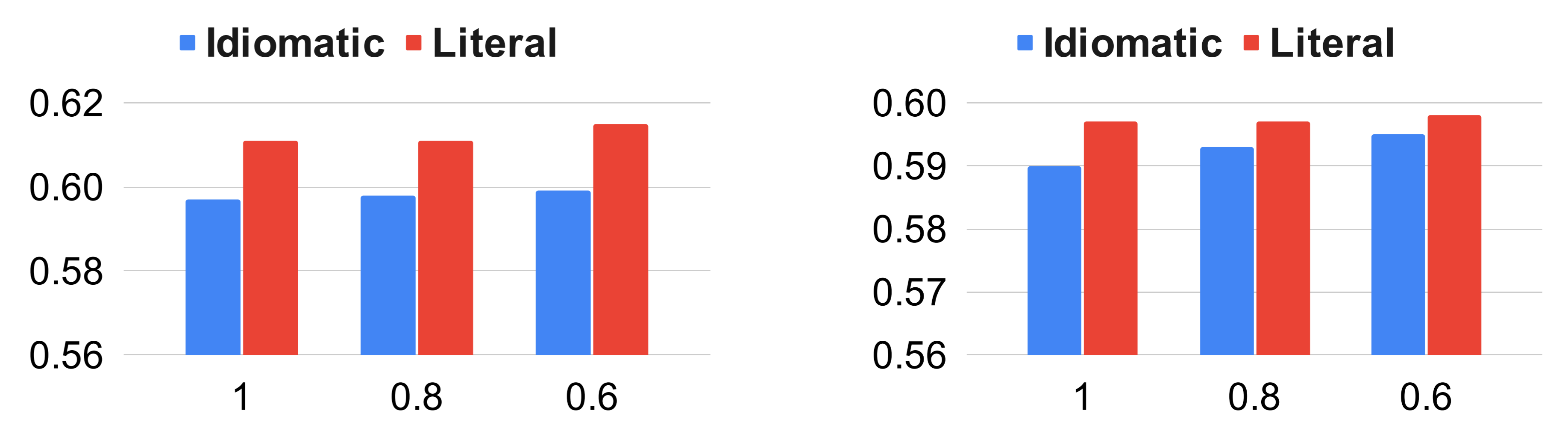| S1 | S2 | S3 |
|---|---|---|
| Fiz isso, inclusive, na exibição do último episódio da série, quando era editor da Rolling Stone. | Li o resumão (era contra até então), fiz um textão completamente desacreditado pelo que virou a minha profissão e de coração partido pelo episódio mequetrefe. | O final era estranhamente confuso, talvez condizente com o que vinha acontecendo na série. |

*A data sample from the PT data. EN samples are not included because of space limit.*

## RESEARCH QUESTION

Do language models trained on extensive text corpora of human language, perform differently or similarly under contexts containing literal and idiomatic expressions?

## OVERALL ARCHITECTURE

Add special token

S1 + S2 → Prompt → Model → Output → Sentence Tokenizer → S3'

## RESULTS & DISCUSSION

### Automatic Evaluation



BERTScore of GPT-3 davinci model under zero-shot settings for EN (left) and PT (right). X-axis: different temperature settings. They shows that literal contexts are easier for LLMs than idiomatic ones by a small factor.

**EN 🇬🇧 vs PT 🇵🇹 Results**
- similar results because of the same language family and both being high resource languages

Zero shot vs Few shot vs Finetuned
- For EN, zero shot was better whereas, finetuning helped the PT data.

### Human Evaluation

Task 1: Relevancy Rating of S3'
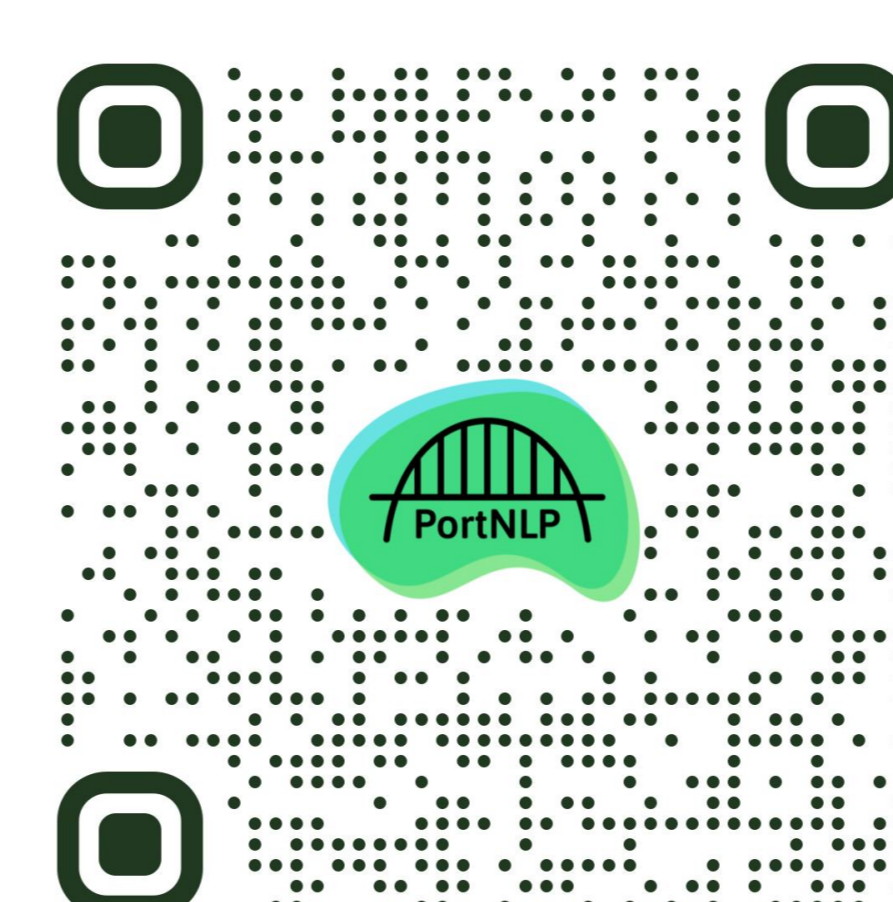GPT-3 results generated better continuations in idiomatic cases.
Task 2: Grammatical correctness (Subjective)
There were incomplete sentences, irrelevant texts, unexpected content/nonsensical contents.

## WHAT NEXT?

- Variation of prompts to be explored.
- Usage of more recent models.
- What about idioms in low resource languages?

## CODE & PAPER

pokharel@pdx.edu
ameeta@pdx.edu