

# *Pun Word Detection, Location, Explanation*

## **Group 1**

**Melissa Kronenberger, Vinicius Lima, Aidan Sommers, Adrienne Zhang**  
**CNIT 51900**

# Getting Started

## Basic Plan

1. Prune Stopwords
2. Collect all word senses of remaining words
  - Evaluate if similarity is relevant for contextual association
3. Provide an LLM with sentence and background information
  - Prompt for reasoning which word and which senses comprise the pun
  - Prompt for neat, step-by-step output in explanation of the pun
  - Fine-tune on a dataset
4. Augment the algorithm with an initial pun detection step

# Finding Tools and Data

## Shopping List

### Word Senses and Simpath

- WordNet through the Natural Language Toolkit (NLTK)

### Large Language Model Usage

- Nous-Hermes-Llama-2 13B
  - Fine-tuned (by us) on pun evaluation input and response
- OpenAssistant LLaMA 2 SFT v10 (70B)
  1. Prompted for a simple yes/no classification
  2. Prompted with examples and unformatted explanations
  3. Prompted with neatly formatted step by step explanation for how to breakdown a pun

### Dataset

- SemEval-2017 Task 7: Detection and Interpretation of English Puns
  - Plentiful related papers
  - Resource dataset contains detection, location, and interpretation data
  - Interpretations use WordNet word senses

## Nous-Hermes-Llama-2 13B Prompt and Fine-tuning Dataset

- Fine-tuned on 1500 examples from SemEval 2017
- Prompt structure for fine-tuning:

### Instruction: Use the context to identify which word in the following sentence is the 'pun word' with the double meaning, it is possible that there is no 'pun word' if there is no word with two meanings in this context. The given sentence is '**A scientist doing a large experiment with liquid chemicals was trying to solve a problem when he fell in and became part of the solution.**'

### Response: The 'pun word' is solution. In the context of the given pun, solution has the two meanings (senses) of:

1. 'a homogeneous mixture of two or more substances; frequently (but not necessarily) a liquid solution'
2. 'a method for solving a problem'

## Nous-Hermes-Llama-2 13B Prompt and Fine-tuning Dataset

### Instruction: Use the context to identify which word in the following sentence is the 'pun word' with the double meaning, it is possible that there is no 'pun word' if there is no word with two meanings in this context. The given sentence is '**A scientist doing a large experiment with liquid chemicals was trying to solve a problem when he fell in and became part of the solution.**'

### Response: The 'pun word' is solution. In the context of the given pun, solution has the two

1. 'a homogeneous mixture of two or more substances; frequently (but not necessarily) a liquid solution'
2. 'a method for solving a problem'

**\*Respective Synsets from WordNet\***

## Nous-Hermes-Llama-2 13B (Fine-tuned) Performance

- Given an input that is pun:
  - Correctly identified the 'pun word', the word with the double meaning, in 90.68% of the examples tested
- Given an input that is not a pun:
  - Can generally distinguish a typical sentence such as "How are you?" as not being a pun
  - Performs very poorly if the given sentence is a phrase or proverb such as "Time is money", and will classify such inputs as puns

# Methods

## OpenAssistant LLaMA 2 SFT v10 (70B) - Simple yes/no for pun vs no pun

```
<|im_start|>system  
{system_message}<|im_end|>  
<|im_start|>user  
{prompt}<|im_end|>  
<|im_start|>assistant
```

# Methods

## OpenAssistant LLaMA 2 SFT v10 (70B) - Simple yes/no for pun vs no pun

```
<|im_start|>system  
{system_message}<|im_end|>  
<|im_start|>user  
{prompt}<|im_end|>  
<|im_start|>assistant
```

**system\_message** = "You are a helpful assistant who will be asked to determine if a given context contains a pun. You will only respond with yes or no"



# Methods

## OpenAssistant LLaMA 2 SFT v10 (70B) - Simple yes/no for pun vs no pun

```
<|im_start|>system  
{system_message}<|im_end|>  
<|im_start|>user  
{prompt}<|im_end|>  
<|im_start|>assistant
```

**system\_message** = "You are a helpful assistant who will be asked to determine if a given context contains a pun. You will only respond with yes or no"

**prompt** = f"A pun is a form of word play that uses multiple interpretations of a word in a given context. The pun word is the word with multiple possible interpretations in the given context. In the given context, the pun word will have a more common interpretation and also an alternative interpretation that relies on a supporting term from a different part of the context. The pun word is more likely to appear in the latter part of the given context, as there is often a setup with the supporting term before the pun word. \n For example: Is this a pun? 'I used to be a banker, but I lost interest.' Simply answer yes or no: yes. \n For example: Is this a pun? 'Time is money.' Simply answer yes or no: no. \n Is this a pun? '{sentence}' Simply answer yes or no: "

## OpenAssistant LLaMA 2 SFT v10 (70B) - Simple yes/no for pun vs no pun

### **prompt:**

A pun is a form of word play that uses multiple interpretations of a word in a given context. The pun word is the word with multiple possible interpretations in the given context. In the given context, the pun word will have a more common interpretation and also an alternative interpretation that relies on a supporting term from a different part of the context. The pun word is more likely to appear in the latter part of the given context, as there is often a setup with the supporting term before the pun word.

For example: Is this a pun? 'I used to be a banker, but I lost interest.' Simply answer yes or no: yes.

For example: Is this a pun? 'Time is money.' Simply answer yes or no: no.

Is this a pun? 'Always trust a glue salesman, because they stick to their word.' Simply answer yes or no:

Huang, Y.-H., Huang, H.-H., & Chen, H.-H. (2017). Identification of homographic pun location for Pun Understanding. *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*.  
<https://doi.org/10.1145/3041021.3054257>

# Results

## OpenAssistant LLaMA 2 SFT v10 (70B) - Simple yes/no for pun vs no pun

- Correctly classifies pun vs no pun for 73.33% of the examples given

## OpenAssistant LLaMA 2 SFT v10 (70B) - Prompt with examples, no format

- Tested this model's performance **not using the system prompt** or a principled prompt structure

## OpenAssistant LLaMA 2 SFT v10 (70B) - Prompt with examples, no format

- Prompt example (if needed for presentation)

A pun is a form of word play that uses multiple interpretations of a word in a given context.

The pun word is the word with multiple possible interpretations in the given context.

In the given context, the pun word will have a more common interpretation and also an alternative interpretation that relies on a supporting term from a different part of the context. The pun word is more likely to appear in the latter part of the given context, as there is often a setup with the supporting term before the pun word.

The following are **three** examples and explanations of puns in different contexts:

Example 1. : 'I used to be a banker, but I lost interest.'

Explanation of 1. : The pun word is 'interest', which has two interpretations in this context: 'a sense of concern or curiosity' and 'a fixed charge for borrowing money'. In this context, 'banker' is the support term for the pun word 'interest'. The first interpretation of 'interest', 'a sense of concern or curiosity', can be understood even if the word banker is replaced with another job title like engineer. In contrast, the second meaning of 'interest', 'a fixed charge for borrowing money', is supported by the term 'banker'.

Example 2. ...

Using similar thinking, identify and explain the pun word in the following context by identifying which word can have multiple interpretations given different parts of the context. Explain the common interpretation in the given context and then use the supporting term to explain the alternative interpretation.

New pun: **'Always trust a glue salesman, because they stick to their word.'**

Explanation of new pun: The pun word is

# Results

## OpenAssistant LLaMA 2 SFT v10 (70B) - Prompt with examples, no format

- Given an input that is pun:
    - Correctly identified the 'pun word', the word with the double meaning, in 85.37% of the examples tested
  - Given an input that is not a pun:
    - Inconsistent output, difficult quantitatively to discern without manual labeling
- Generally performed with similar limitations to Nous-Hermes-Llama-2 13B on phrases and proverbs

# Methods

## OpenAssistant LLaMA 2 SFT v10 (70B) - Prompted with neatly formatted step by step explanation for how to breakdown puns

- Tested this model's performance **using the system prompt** and Chain-of-Thought Prompting

Wei, J., Zhou, D., Le, Q., Chi, E., Xia, F., Ichter, B., Bosma, M., Schuurmans, D., & Wang, X. (2022). *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*.  
<https://doi.org/https://doi.org/10.48550/arXiv.2201.11903>

# Methods

## OpenAssistant LLaMA 2 SFT v10 (70B) - Prompted with neatly formatted step by step explanation for how to breakdown puns

```
<|im_start|>system  
{system_message}<|im_end|>  
<|im_start|>user  
{prompt}<|im_end|>  
<|im_start|>assistant
```

system\_message = "You are a helpful assistant who will be asked to determine if a given context contains a pun. You will think step by step by writing your steps in determining if a given context contains a pun. After your step by step response, simply say 'my final answer is no' or 'my final answer is yes'."

prompt = f"A pun is a form of word play that uses multiple interpretations of a word in a given context. The pun word is the word with multiple possible interpretations in the given context. In the given context, the pun word will have a more common interpretation and also an alternative interpretation that relies on a supporting term from a different part of the context. The pun word is more likely to appear in the latter part of the given context, as there is often a setup with the supporting term before the pun word. Now here are some examples of thinking through the questions step by step: \n For example: Is this a pun? 'I used to be a banker, but I lost interest.' Explanation: The context can be broken into two parts of 1. I used to be a banker and 2. but I lost interest. The word that has two interpretations in this context is interest. The interpretation of interest related to part 2. is 'a sense of concern or curiosity'. The interpretation of interest related to part 1. is in relation to the term banking, in this part interest refers to 'a fixed charge for borrowing money'. So, because there is a word with multiple interpretations related to different parts of the given context, this is a pun. My final answer is yes. \n For example: Is this a pun? 'Time is money.' There is no word in this context that can have multiple interpretations, so this is not a pun. My final answer is no. \n Now it is time for the assistant to help: \n Is this a pun? '{sentence}' Let's work this out in a step by step way to be sure we have the right answer."



# Methods

## OpenAssistant LLaMA 2 SFT v10 (70B) - Prompted with neatly formatted step by step explanation for how to breakdown puns

- Prompt example (if needed for presentation)

A pun is a form of word play that uses multiple interpretations of a word in a given context. The pun word is the word with multiple possible interpretations in the given context. In the given context, the pun word will have a more common interpretation and also an alternative interpretation that relies on a supporting term from a different part of the context. The pun word is more likely to appear in the latter part of the given context, as there is often a setup with the supporting term before the pun word.

Now here are some examples of thinking through the questions step by step:

For example: Is this a pun? 'I used to be a banker, but I lost interest.' Explanation: The context can be broken into two parts of 1. I used to be a banker and 2. but I lost interest. The word that has two interpretations in this context is interest. The interpretation of interest related to part 2. is 'a sense of concern or curiosity'. The interpretation of interest related to part 1. is in relation to the term banking, in this part interest refers to 'a fixed charge for borrowing money'. So, because there is a word with multiple interpretations related to different parts of the given context, this is a pun. My final answer is yes.

For example: Is this a pun? 'Time is money.' There is no word in this context that can have multiple interpretations, so this is not a pun. My final answer is no.

Now it is time for the assistant to help: \n Is this a pun? '{sentence}' Let's work this out in a step by step way to be sure we have the right answer."

# Methods

**OpenAssistant LLaMA 2 SFT v10 (70B) - Prompted with neatly formatted step by step explanation for how to breakdown puns**

- Prompt example (if needed for presentation) - continued

For example: Is this a pun? 'I used to be a banker, but I lost interest.'

Explanation:

The context can be broken into two parts of 1. I used to be a banker and 2. but I lost interest.

The word that has two interpretations in this context is interest. The interpretation of interest related to part 2. is 'a sense of concern or curiosity'. The interpretation of interest related to part 1. is in relation to the term banking, in this part interest refers to 'a fixed charge for borrowing money'.

So, because there is a word with multiple interpretations related to different parts of the given context, this is a pun. My final answer is yes.

# Results

## OpenAssistant LLaMA 2 SFT v10 (70B) - Prompted with neatly formatted step by step explanation for how to breakdown puns

- Correctly classifies pun vs no pun for 74.26% of the examples given

This final classification is sometimes not aligned well with the explanation given.

The explanation given may be sufficient, but the final answer may be classified differently due to some ambiguity in the model's understanding of what to consider a pun

- Substantially improved results for explanation

Further fine-tuning on step by step output and the final pun classification would be expected to greatly improve results.

- See the proposed “Evol-Instruct” method in <https://arxiv.org/abs/2304.12244>

# Results

## OpenAssistant LLaMA 2 SFT v10 (70B) - Prompted with neatly formatted step by step explanation for how to breakdown puns

- Example step by step output:

1. First, let's identify the different parts of the given context: **"There was a contest on a slippery rock slope."** and **"I won in a landslide."**
2. Next, we look for a word with multiple interpretations in these parts of the context:
  - The word "landslide" could have two interpretations. In the first part, it might refer to a literal landslide caused by the slippery rock slope. In the second part, it could refer to winning the contest by a large margin.
3. We now need to check if the word "landslide" has a supporting term from another part of the context:
  - The supporting term in this case is "contest," which helps us understand that "landslide" refers to winning the contest by a large margin in the second part of the context.
4. Since the word "landslide" has multiple interpretations related to different parts of the context and has a supporting term, we can conclude that this is indeed a pun.
5. My final answer is yes.

# Limitations

- Methodological limitations
  - Designed for Homographic puns with a single “pun word” and a supporting term in another part of the context
  - Reliant on LLM’s Black Box “Reasoning” Capabilities
    - Responsible for making associations based on given pun
- Technical limitations
  - Fine tuning of OpenAssistant LLaMA 2 SFT v10 (70B) would be preferable
    - Not feasible for this project due to time constraints and compute constraints
    - Would create a dataset of step-by-step reasoning-based breakdowns of puns
    - Would also QLORA fine-tune for more structured output

# Addressing Weaknesses - Augmentations

## Improving Pun Detection/Sentence Classification

### Two Strategies for Pun Yes/No Detection

1. Long Short-Term Memory (LSTM) Classification
2. Support Vector Machine (SVM) Text Classification

### More Puns to Train On

- Reddit's "Humor Detection" Puns, by Orion Weller

### Proverbs to Train On

- "Proverb Database: Corpus of American English Proverbs (CAEP) and Experimental Study," by Lückert, Claudia

### Balance Puns/Non-Puns to 50/50% to Remove Bias

# Augmentations

## Pun Classification: Alternate Approach #1

### LSTM Classification Algorithm

- Model with raw data (SemEval 2017): unbalanced data.
- 128 LSTM units, 0.2 dropout, Adam optimizer, 20% testing data, 5 epochs..

Puns	Non-puns
2878	1152

### Results

- 84% accuracy on validation dataset.
- Incorrectly detected "The crane lifted a heavy load."

```
# New sentence to test
```

```
new_sentence = "The crane lifted a heavy load."
```

```
1/1 [=====] - 0s 345ms/step
```

```
Predicted probabilities: [[3.6190318e-05 9.9996376e-01]]
```

```
Predicted class: [1]
```

# Augmentations

## Pun Classification: Alternate Approach #1

### LSTM Classification Algorithm

- New approach: balancing the data: undersampling.
- 128 LSTM units, 0.2 dropout, Adam optimizer, 20% testing data, 5 epochs.

Puns	Non-puns
1152	1152

### Results

- 82% accuracy on validation dataset.
- Correctly detected "The crane lifted a heavy load."

```
# New sentence to test
```

```
new_sentence = "The crane lifted a heavy load."
```

```
1/1 [=====] - 0s 42ms/step
```

```
Predicted probabilities: [[0.99397814 0.00602187]]
```

```
Predicted class: [0]
```



# Augmentations

## Pun Classification: Alternate Approach #2

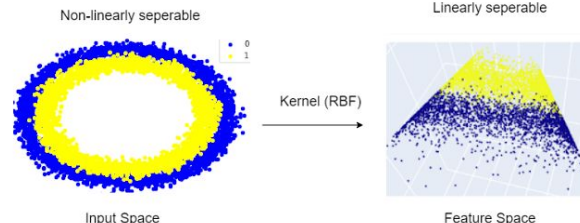
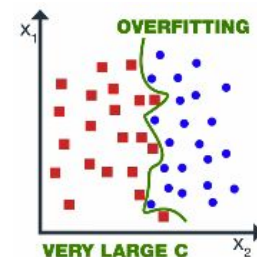
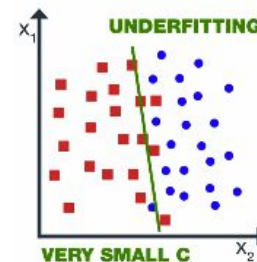
### SVM Text Classification

- Focused on higher precision score
  - Goal: reduce number of false pun classifications (false positives)
- Vectorization of inputs using bag-of-words approach
- Parameters: {C: 10, gamma: 0.1, kernel: 'rbf'}
  - C - determines error margin of decision line
  - gamma - determines curvature level of decision line
  - Radial Basis Function (rbf) - useful for non-linear decision boundaries

### Results

- Accuracy score: 85.54%
- Precision scores
  - Classifying puns: 80%
  - Classifying non-puns: 87%

Biased towards non-pun classification as dataset has more non-puns and proverbs than puns



# DEMO



# Questions?

[polytechnic.purdue.edu](https://polytechnic.purdue.edu)



Begin Leftover Slides from Adrienne's Original Skeleton

# Project Description (delete later)

Your task is to detect sentences/paragraphs that use multiple meanings of words or phrases. Jokes, puns, riddles are a perfect examples of it -- feel free to model them or use them.

To detect a word that is used in multiple meanings, your program will need to identify the word and outline the used meanings. A user will enter a text and the program will either have to identify a word (or part of a word) that can be used in more than one sense in this context (and its meanings) or declare that no such word exists in this text. The meanings have to be compatible with at least some portions of the text.

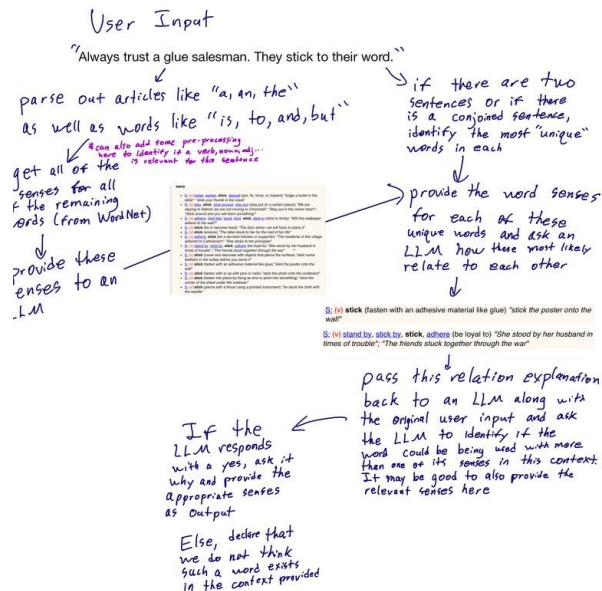
Do not work with words that sound similar -- you will do this later in the semester, as of now we did not count sounds yet. You can, however, use part of a word as one of the senses if necessary, or break one word into two and use the phrase as the second meaning (see Example 2) if its meaning is different from the meaning of the original word.

As you can see, this project is rather open-ended: the goal is for you to play with multiple meanings of words, and as long as you achieve it (minus hard-coding), you will get full credit. It is not expected that the program will work perfectly. You may elect to give hints to your program if it stumbles, but please be reasonable in this case :) If you do give hints, please explain in your write-up why it was needed, what didn't work as expected, etc.

You are to submit a write-up of the project (in English, no code, although feel free to use examples) and a presentation that you used to showcase the project in class. I may request code as well from some teams, however, it's the analysis of the program that will be evaluated, not your ability to code.

# Overview

## Overall approach/plan →



\* Augmentations, tuning, and further layers can be used at various points in order to improve performance

\* We may also use a couple of different methods and attempt to aggregate these for a final response

# Tools & Technologies

Wordnet

NLTK

etc.