



CMPT 413/713: Natural Language Processing

Project information

Spring 2024
2024-02-14

Summary of topics

- So far
 - Language Modeling: $P(w_i | w_{1:i-1})$
 - Text classification: $P(c | d)$
 - with Naive Bayes, Logistic Regression and Neural Networks
 - Static and contextual word embeddings: Representing w as a vector
 - Sequence modeling with RNNs and Transformers
- After the break
 - Focus on project
 - More on pretraining and fine-tuning LLMs
 - Structured representations and parsing
 - Other NLP applications and using LLMs

Project

- Project timeline
 - 2-22 - Initial abstract/title (not graded) for initial feedback
 - 2-29 - Proposal due
 - 3-21 - Project milestone due
 - 4-6 - Final project videos
 - 4-10 - Final project report due

Project Proposal

- Initial title and abstract (~250 words)
 - Think about what you want to do for your project
 - Have an informative title (“CMPT 413 class project” is not a good title)
 - Describe the topic/problem your group will work on
 - Scope of the project
 - Data and compute resources you plan to use
- Proposal (1-2 pages)

<https://angelxuanchang.github.io/nlp-class/project.html>

Project Proposal

- What task are you addressing? What is the input / output? Why is it interesting?
- What specific aspects will your project be on?
 - Re-implement paper? Compare different methods? Analysis?
- What **data** do you plan to use? For training vs for evaluation?
 - Preliminary statistics for your data (number of sentences, tokens, etc)
- What is the specific method or methods you will use to address the task?
 - **What will you implement by yourself vs what existing code will you use?**
 - What **compute resources** do you plan to use?
- How do you plan to evaluate?
 - Data splits?
 - What **metrics**?
 - What experiments will you run to **compare** different variations / different approaches?
- **Timeline and work breakdown**
 - What do you plan to have by the milestone? The end of the term?
 - Who will work on what?

What kind of approaches you want to take

- In this class, we started with fundamentals and simple ML models (Naive Bayes and Logistic regression)
- You should be experimenting with methods beyond that
 - Use pretrained models! Do fine-tuning!
 - Use LLMs with few-shot prompting
 - We will cover more about these topics after the break

Type of problem and resources

<https://angelxuanchang.github.io/nlp-class/project.html>

- Pick a problem based on what are you interested in?
- Types of problems
 - Building a cool NLP application
 - Re-implementing / re-reproducing a recent paper
 - Applying an existing neural model to a new task
 - Implementing a complex neural architecture
 - Proposing a new neural model or a new variation of an existing model
 - Proposing a new training, optimization, or evaluation scheme
 - Experimental and/or theoretical analysis of a NLP model
 - Probing the capabilities of current NLP models

Some NLP Tasks

- SQuAD: question answering
- SNLI: natural language inference
- SRL: semantic role labeling
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis
- POS: Part-of-speech tagging
- Entity linking
- Constituency Parsing
- Dependency Parsing
- Intent detection and Slot filling
- Machine translation
- Paraphrasing
- Summarization
- Semantic Parsing

Tasks

- SQuAD: question answering
- SNLI: natural language inference
- SRL: semantic role labeling
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

Stanford CS224 default project

– SQuAD

– RobustQA

Passage

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

Question / Answer

What causes precipitation to fall?

gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?

graupel

Where do water droplets collide with ice crystals to form precipitation?

within a cloud

SQuAD: 100,000+ Questions for Machine Comprehension of Text

Tasks

- SQuAD: question answering
- SNLI: natural language inference
- SRL: semantic role labeling
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

Premise

Hypothesis

contradiction

entailment

neutral

A man inspects the uniform of a figure in some East Asian country.

The man is sleeping.

An older and younger man smiling.

Two men are smiling and laughing at the cats playing on the floor.

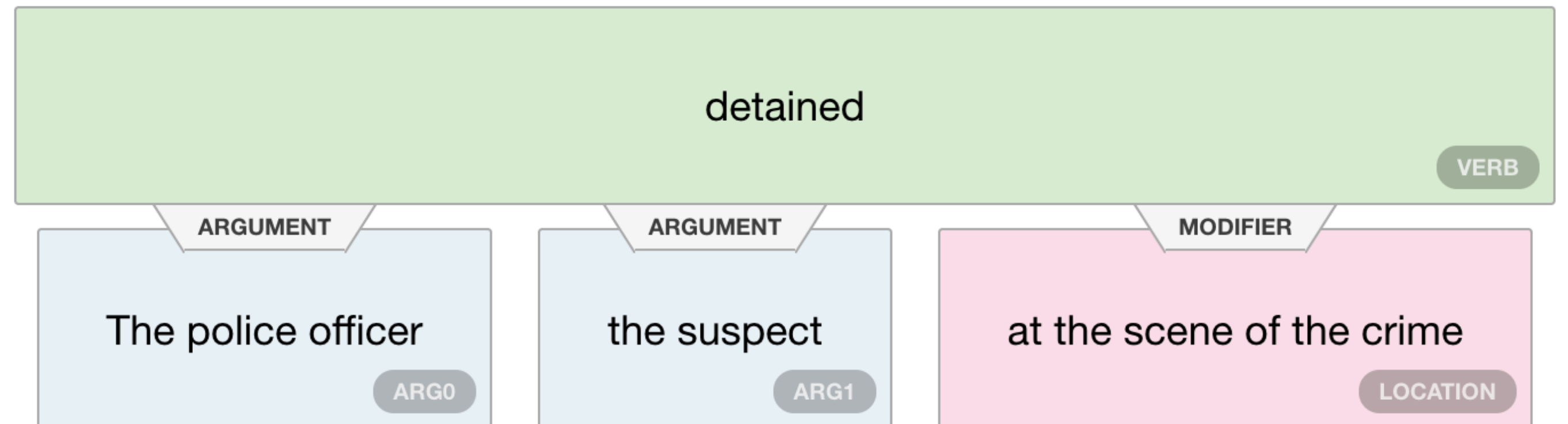
A soccer game with multiple males playing.

Some men are playing a sport.

Tasks

- SQuAD: question answering
 - SNLI: natural language inference
 - **SRL: semantic role labeling**
 - Coref: coreference resolution
 - NER: named entity recognition
 - SST-5: sentiment analysis
-
- Different types of frames
 - FrameNet vs PropBank
 - Different semantics roles depending on the verb

The police officer detained the suspect at the scene of the crime

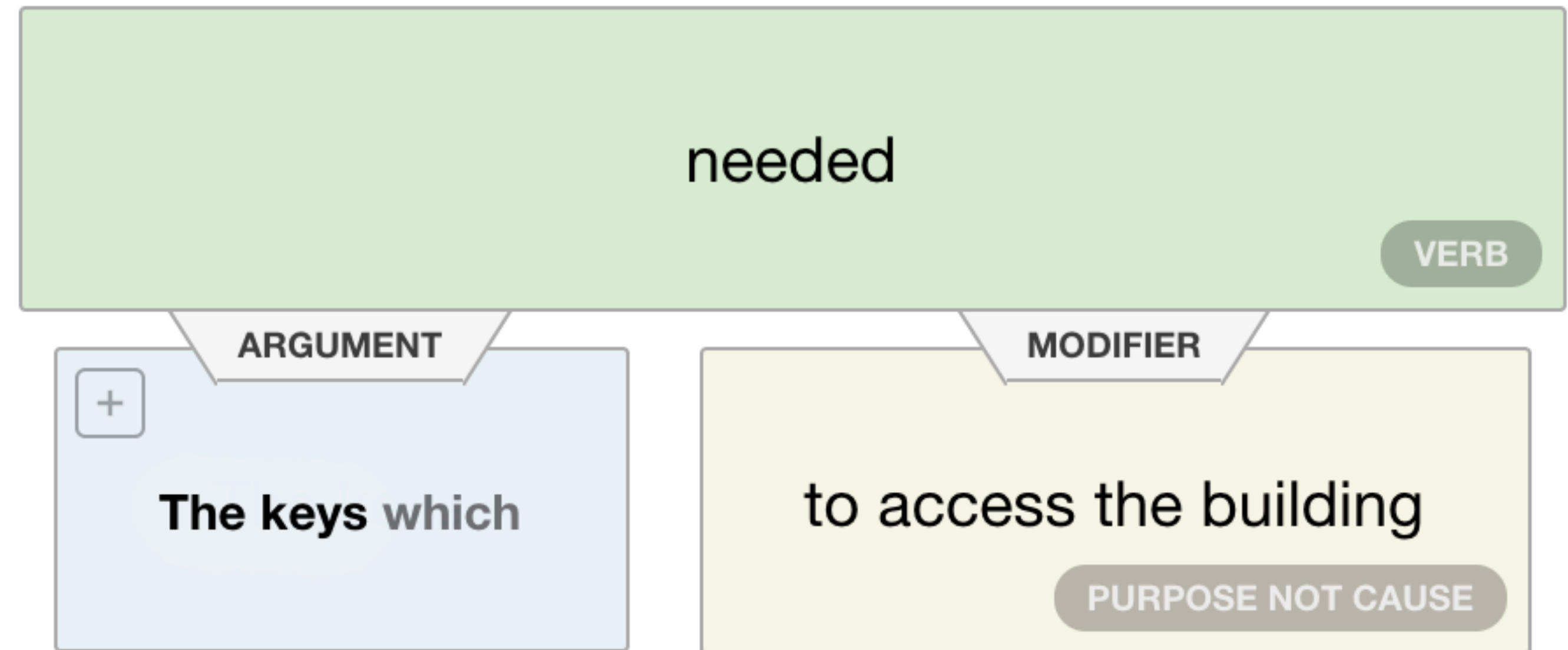


(figure credit: <https://demo.allennlp.org/semantic-role-labeling/>)

Tasks

- SQuAD: question answering
- SNLI: natural language inference
- **SRL: semantic role labeling**
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

The **keys**, which were **needed** to access the building, were locked in the car.



(figure credit: <https://demo.allennlp.org/semantic-role-labeling/>)

Tasks

- SQuAD: question answering
- SNLI: natural language inference
- **SRL: semantic role labeling**
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

The keys, which were needed to **access** the building, were locked in the car.

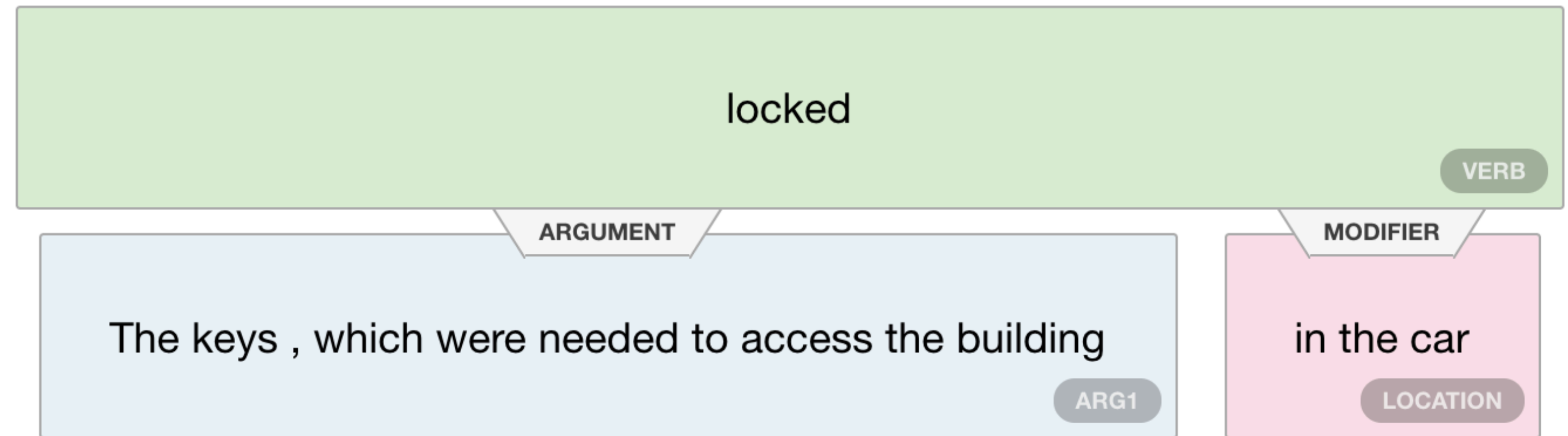


(figure credit: <https://demo.allennlp.org/semantic-role-labeling/>)

Tasks

- SQuAD: question answering
- SNLI: natural language inference
- **SRL: semantic role labeling**
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

The keys, which were needed to access the building, were **locked** in the car.



(figure credit: <https://demo.allennlp.org/semantic-role-labeling/>)

Tasks

- SQuAD: question answering
- SNLI: natural language inference
- SRL: semantic role labeling
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

*“I voted for **Nader** because **he** was most aligned with **my** values,” **she** said.*

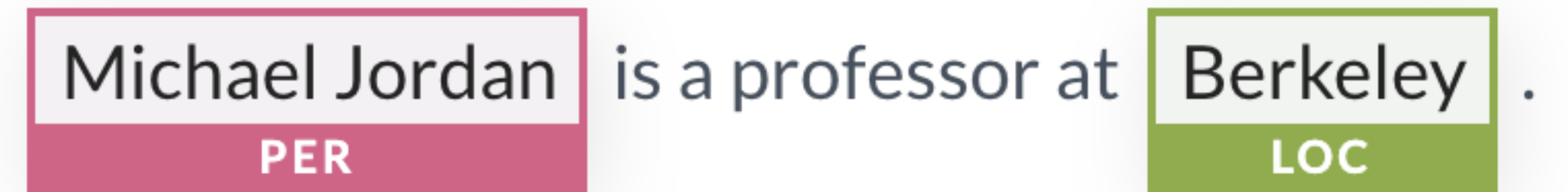
The diagram shows three curved arrows indicating coreference relations: one from 'Nader' to 'he', one from 'my' to 'she', and one from 'I' to 'she'.

(figure credit: <https://nlp.stanford.edu/projects/coref.shtml>)

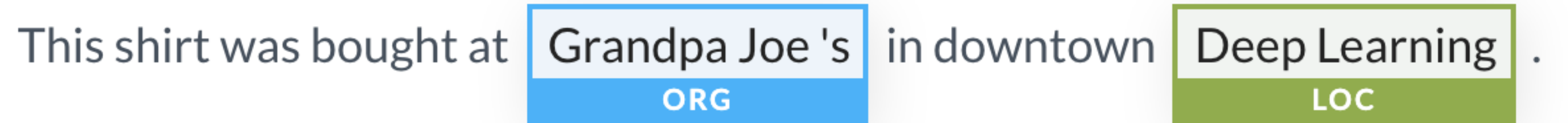
Tasks

- SQuAD: question answering
- SNLI: natural language inference
- SRL: semantic role labeling
- Coref: coreference resolution
- **NER: named entity recognition**
- SST-5: sentiment analysis

Michael Jordan is a professor at Berkeley .



This shirt was bought at Grandpa Joe 's in downtown Deep Learning .



Did Uriah honestly think he could beat The Legend of Zelda in under three hours ?



(figure credit: <https://demo.allennlp.org/named-entity-recognition>)

Tasks

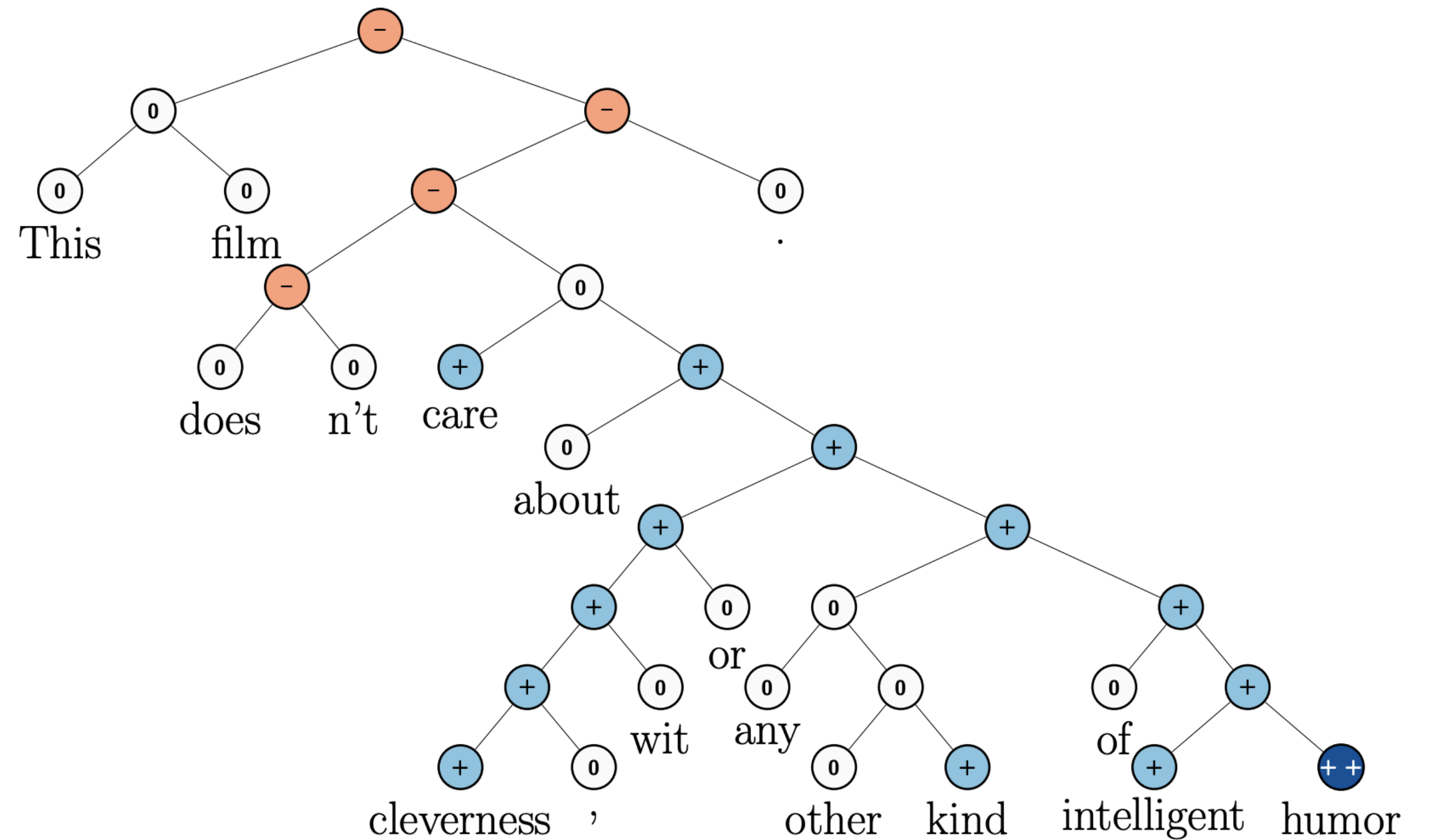
- SQuAD: question answering
- SNLI: natural language inference
- SRL: semantic role labeling
- Coref: coreference resolution
- NER: named entity recognition
- SST-5: sentiment analysis

5 classes

- Very Positive
- Positive
- Neutral
- Negative
- Very Negative

Stanford Sentiment Treebank

This file doesn't care about cleverness, wit or any other kind of intelligent humor.



Recursive deep models for semantic compositionality over a sentiment treebank