QUESTION ANSWERING

Dataset & Input/Output

- 120 **Multiple Choice** Questions from AI2 dataset + decomposed answers
- Questions taken from **4th grade science exam** NYSE REGENTS
- 1200 sentences from Barron's corpus

Input is the question with 4 options

If an object is attracted to a magnet, it is most likely made of

a) wood b) plastic c) cardboard d) metal

Output is the predicted answer

d) metal

Tools

- NLTK (Python)
- WordNet
- FrameNet
- SEMAFOR
- Porter’s Stemmer

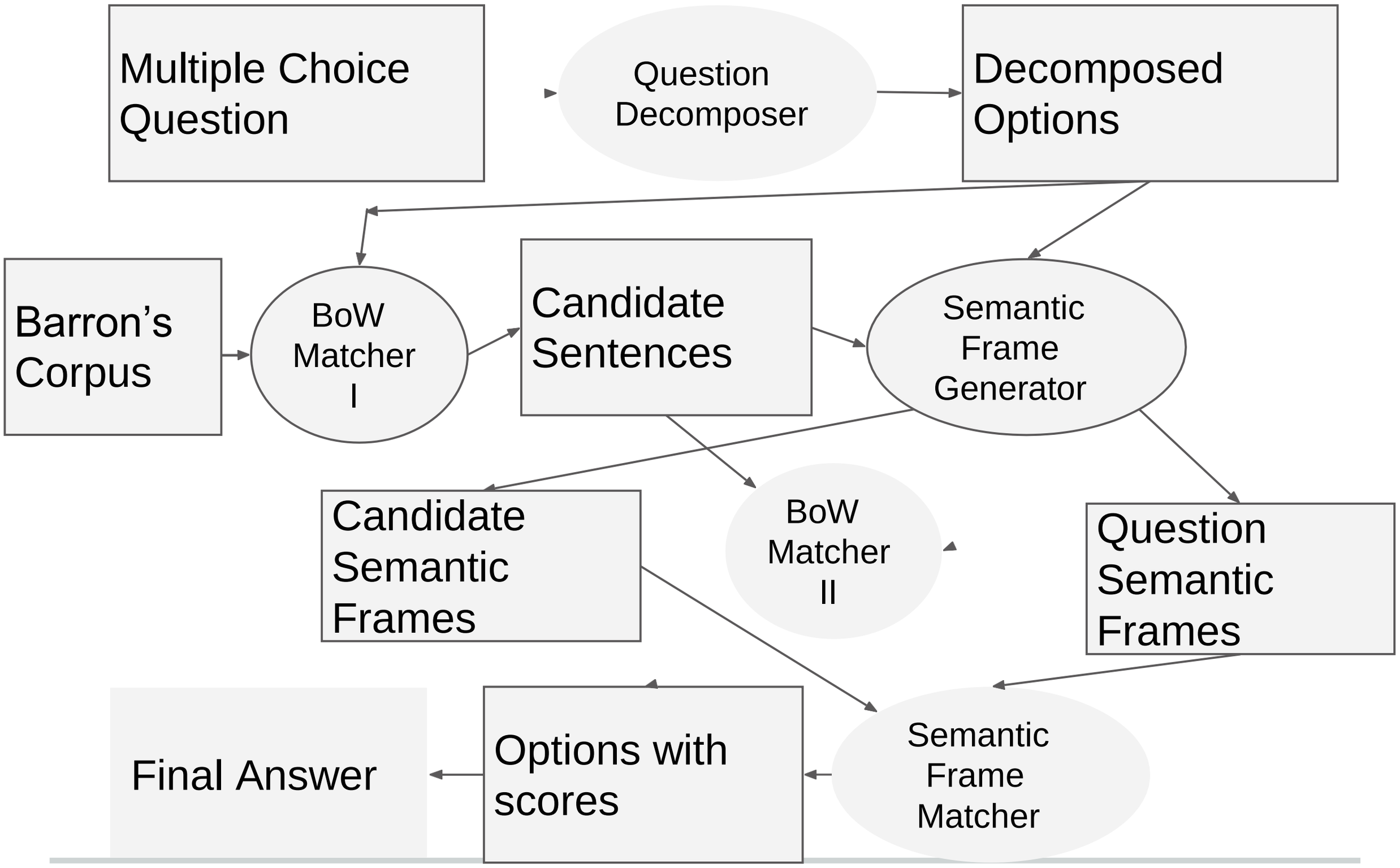
Key Components

Two main components to the Question Answering task

- Extraction of the candidate sentences from the corpus

- Answer Scoring

Framework


Multiple Choice Question
Question Decomposer
Decomposed Options
Barron's Corpus
BoW Matcher I
Candidate Sentences
Semantic Frame Generator
Candidate Semantic Frames
BoW Matcher II
Question Semantic Frames
Final Answer
Options with scores
Semantic Frame Matcher

Methods

Candidate Extraction (I)	Answer Scoring (II)
Bag of Words(BoW)	Bag of Words
Semantic Frame Matching(SFM)	Semantic Frame Matching
Bag of Words	Semantic Frame Matching
Bag of Words	Weighted BoW + SFM
Bag of Words	BoW/SFM based on confidence
BoW using FOCUS WORDS	SFM using FOCUS WORDS

BoW (I) BoW (II)

BoW(II) : Used Answers in BoW representation and assigned a *score* for each by calculating synset similarities.

BoW(I) : 2 ways of Candidate Extraction

- ➢ Question in BoW representation

Accuracy : 43.7%

- ➢ Union of Candidate Sentences using Answer Sentences in BoW representation

Accuracy : 45.3%

Examples

When plants and animals die, which organisms help return nutrients to the food chain ?

a. decomposers b. predators c. prey d. producers

When options are not being considered, often, irrelevant ***candidate sentences*** are extracted.

Living *things need to take in* ***nutrients*** *in the form of food so that they can grow and create energy.*	***Decomposers*** *are living things that break down dead organisms and recycle their* ***nutrients*** *into the soil.*
Plants *and* ***animals*** *require air, water, light, and nutrients in order to live and survive.*	*Plants are called* ***producers*** *because they provide the* ***food*** *supply for themselves and animals.*

BoW with Focus on Answer options for Candidate Extraction

When comparing each option with candidates, the words in the options were weighted higher than others, which lead to an improvement in the accuracy.

Accuracy : 46.28%

Eg:- Which human activity most often has the most harmful effect on the environment ?

a. breathing b. growing c. planting d. polluting

Harm *to the* ***environment*** *can be caused when animals are displaced and the landscape is changed.*	*Building shelters and the various forms of transportation are some of the ways humans have created* ***pollution.***
Human *decisions and activities have had a major impact on the physical and living* ***environments.***	***Plants*** *and animals depend on each other and the nonliving* ***environment*** *for survival.*

Performance of BoW

- Lack of a proper candidate sentence in the corpus.

Ex : Which food is a fruit ?

a. potato b. onion c. carrot d. pumpkin

➢ Predicts **potato** :(

- Candidate sentence present but meaning is not captured

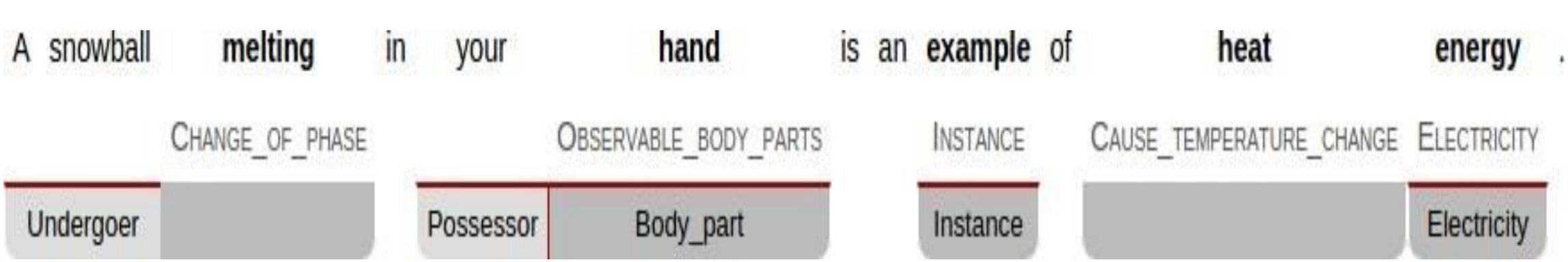
Ex : Which unit of measurement can be used to describe the length of a desk ?

a. degrees celsius b. grams c. litres d. centimeter

We can describe the paper in terms of its color (white) and its **size** (centimeter)

SEMANTIC FRAME MATCHING

Preprocessing - Generated the semantic frame role labels for the entire corpus and Q&A sentences



Three different approaches tried using Semantic Frames.

SEMANTIC FRAME MATCHING

- BoW without FOCUS words as candidate sentences Plus Semantic Frames
  - Accuracy: **51.58%**

- BoW with FOCUS words as candidates Plus Semantic Frames
  - Accuracy: **53.67%**

- Plus FOCUS on semantic frames + other heuristics
  - Accuracy: **59.12%**

Improvements and False Positives

BoW without FOCUS + Semantic Frames

Better Answer matches but wrong candidate sentences lead to wrong answers

Eg:- Which human activity most often has the most harmful effect on the environment ?

a. breathing b. growing c. planting d. polluting

Harm to the environment can be caused when animals are displaced and the landscape is changed.	*Building shelters, using energy, manufacturing goods, and the various forms of transportation are some of the ways humans have created pollution.*
Human decisions and activities have had a major impact on the physical and living environments.	*Plants and animals depend on each other and the nonliving environment for survival.*

Improvements and False Positives

- BoW with FOCUS + Semantic Frames without FOCUS
  - Better candidate sentences but accuracy affected due to noise.

Ex:- Growing thicker fur in winter helps some animals to

a. hide from danger b. attract a mate c. find food

d. keep warm

Candidates:

Animals grow thicker fur during winter season

A polar bear’s fur keeps it warm in extreme temperatures.

Improvements and False Positives

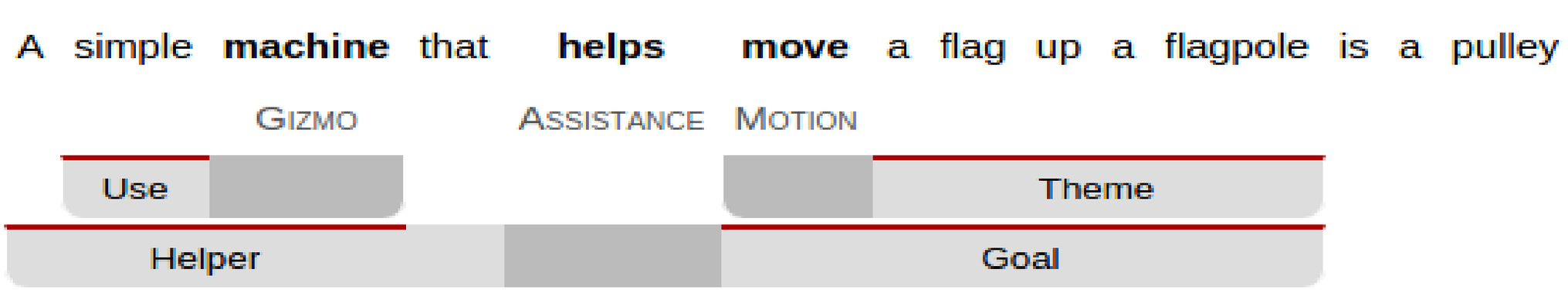
- BoW with FOCUS + Semantic Frames with FOCUS + heuristics
  - Better candidate sentences
  - Much lesser noise
    - But inaccurate semantic role labels!!
  - Joint inference from multiple candidate sentences is still a problem

Improvements and False Positives

Eg:- A simple machine that helps move a flag up a flagpole is a

(a) lever (b) pulley (c) inclined plane (d) bar magnet

Candidate - A pulley is used to raise the flag up the flagpole at your school each day.



Weighted BoW + Semantic Frames can fix this particular scenario (but might break others!)

Improvements and False Positives

Joint inference problem

Eg:- In New York State , the longest period of daylight occurs during

(a) September (b) March (c) December (d) June

Evidence 1 - The longest period of daylight hours occurs at the beginning of Summer.

Evidence 2 - Summer lasts from June 20 to September 21.

Semantic Frames with Confidence Measure

- Incorporated a measure of confidence in the **best answer score**.
  - Tried measures such as general variance, variance from the maximum score etc
  - Switch between **Semantic** and **BoW** scores based on the measure

- Didn’t result in a big change in the accuracy.
  - Helpful sometimes. Not so much otherwise.

Results Summary

Candidate Sentence Extraction Method	Answer Scoring Method	Accuracy
BoW without focus	BoW without focus	45.3%
BoW with focus	BoW without focus	46.28%
BoW with focus	Semantic Frames without focus	53.67%
BoW with focus	Semantic Frames with focus + other heuristics	59.12%

Future Scope

- A better semantic role labeling would greatly help
  - Around 70% of the errors were due to wrong/no semantic frames.
  - Rest of the errors are mostly due to either lack of evidence or the joint inference problem.
- Devise a solution for the joint inference problem
- Incorporate sentiment? (not inherited vs inherited)

Questions?

(Only 4th graders please :P)