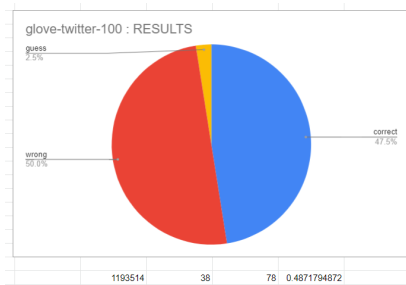


# COMP 472 MP3 - Results Breakdown

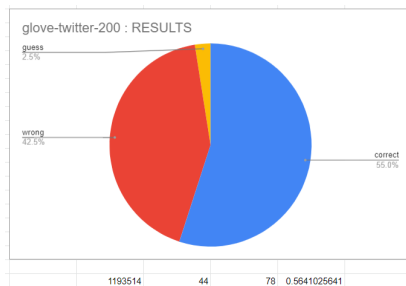
## Models Comparison

### glove-twitter-100 (vocab size 1193514)



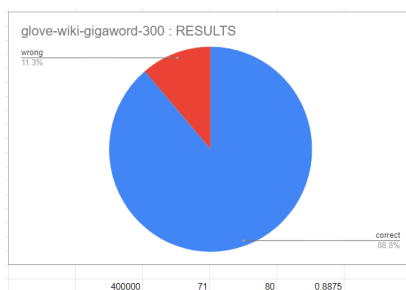
With the smallest embedding size among all the models at 100, this model proved to be the weakest overall, scoring **36.85% below** the human gold standard in accuracy.

### glove-twitter-200 (vocab size 1193514)



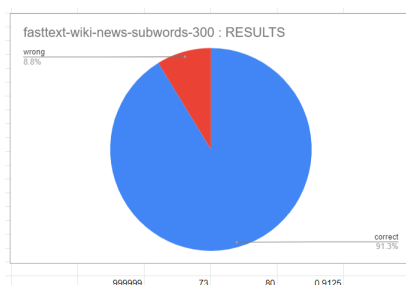
With a middle-ground embedding size of 200, this model was only slightly more potent than its size 100 counterpart, scoring **29.16% below** the human gold standard in accuracy.

### glove-wiki-gigaword-300 (vocab size 400000)



Now at the largest embedding size of 300, we have a significantly more powerful model, scoring **3.18% above** the human gold standard in accuracy.

### fasttext-wiki-news-subwords-300 (vocab size 999999)



Using the larger of the two corpora at size 300 embedding, we have the most powerful model of all the experiments, scoring **5.68% above** the human gold standard in accuracy.

# Notable Differences and Speculation

## Against a Random Baseline

Given a purely random choice between 4 options, with only 1 correct answer within the set, the average rate of success will **tend towards 25% accuracy** once the data set is sufficiently large. As such, **all of the above models are significantly more effective** (positive accuracy delta in excess of 23.72% for all models) when compared to the random baseline.

## Performance Between Models

There is a notable increase in accuracy between the models based on two factors:

- An **increase** in word embedding/vector size;
- An **increase** in vocabulary size.

It should be noted that, based on the data observed from these experiments, these increases are **not strictly linear** in scale, **nor equally weighted** in importance:

- In the case of embedding changes with a static vocabulary using *glove-twitter*, increasing the embedding size by 100 (from 100 to 200) raised the accuracy by 7.69%. However, **increasing the embedding size again from 200 to 300** (given a new dataset *glove-wiki-gigaword*, which may have also had an effect) **raised the accuracy by 32.34%, a relative increase of over 4x**.
- In the case of vocabulary changes with a static embedding using *glove-wiki-gigaword-300* and *fasttext-wiki-news-subwords-300*, increasing the vocabulary size by 599999 (from 400000 to 999999) raised the accuracy by 2.5%. However, an increase of 793,514 (from 400000 to 1193514) was not enough to increase the accuracy significantly when changing from a larger embedding size to a smaller one (300 to 200, across models). **This implies that the weight of the vocabulary size relative to the embedding size, given a large enough dataset of words, is significantly smaller.**

While more experiments would be necessary to analyze the differences between specific models (glove vs word2vec vs fasttext...), seeing these differences with the parameters is a sign that some parameters (in our case, word embedding size) have a larger effect on the outcome of predictions than others.