Quantifying Ambiguity in Mathematical Word Problems in French and English

Introduction

The main aim of the project was to create a model using entropy. This model would represent a person misinterpreting a question and concluding an answer unexpected by the test maker. The project was prompted by a paper which asked test subjects a word problem which they claimed condensed to a single subtraction such as 15-7=8. The questions used were ambiguous around 70% incorrectly answered the most ambiguous question.

While we could all agree upon inspection of the problems that some were more ambiguous than others, we had no measure to use. Entropy was used as the measuring tool as it has use in different outcomes. Every ambiguous statement introduces a new potential reality to th e problem. The entropy of the system in our model was assessed based on the complete "tree", (tree being a map of the branches of different realities). One reason for this was to mimic the way in which students tackle the problem, reading all the information and then attempting to answer.

"When masters of abstraction run into a concrete wall"**

This paper took 105 university level students and gave them some 6 line long maths questions to solve.

The conclusion of those researchers was that the questions fell in to two categories, Cardinal problems (CP) and Ordinal problems (OP). They claimed that when faced with a CP the students attempted to use a 3 step method, and a 2 step method for OP. They based the questions surrounding a simple calculation such as 15-4=11 but removed the information for the 3 step method so that only the 2 step method was possible. They then asked students to identify when questions were not solvable.

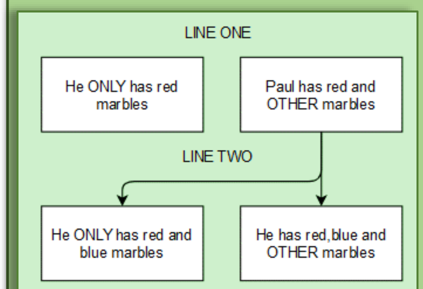
There interpretation of the results was that all test subjects were more likely to say a problem was insolvable when it was a CP rather than OP. Study 1 (S1) was comprised of 85 non-experts and Study 2 (S2), 25 experts. The researchers found that both experts and non-experts were susceptible to thinking "incorrectly" the questions were impossible.

When looking at the questions it appeared that despite the papers claims the only difference between the CP and OP was the type of value, there were other differences. The research project set out to find a way to measure a value for ambiguity using entropy mathematics.

Ordinal and Cardinal Problems

LINE ONE	LINE ONE
Paul has a certain amount of red marbles	Slouchy Smurf is a certain height.
LINE TWO	**LINE TWO**
He also has blue marbles.	He climbs on a Smurf table
LINE THREE	**LINE THREE**
In total, Paul has 14 marbles.	He now attains the height of 14cm.

On the right is the first three lines of two problems from the study. The marble problem is classified as a CP and the height is an OP. The original study says that these sentences are equivalent however an ambiguity exists in lines one and two of the marble problem which is not seen in the height problem.



After the first line there exist two branches, either Paul only has red marbles, or he has red marbles and another colour. The branch which assumed only red is destroyed when the second line states he has red and blue. However, the third line does nothing to eliminate the possibility of a third colour of marble.

For the height question the Smurf's cannot have extra height and as they are stood on the "same table" there is no ambiguity.

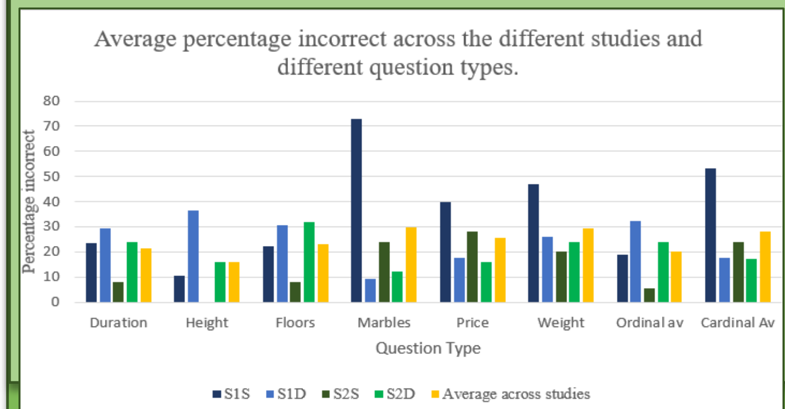
LINE FOUR	LINE FOUR
Jolene has as many blue marbles as Paul, and some green marbles.	Grouchy Smurf climbs on the same table as Slouchy Smurf.
LINE FIVE	**LINE FIVE**
She has two green marbles less than Paul has red marbles.	Grouchy Smurf is 2cm shorter than Slouchy Smurf.
LINE SIX	**LINE SIX**
How many marbles does Jolene have?	What height does Grouchy Smurf attain when he climbs on the table?

However, in LINE 4 of the marble problem it only states "as many as" which could mean Jolene has more blue marbles than Paul. Even if this was changed to the "same as" it is still unclear due to the ambiguity in LINE 2 whether Paul has other marbles affecting the total.

When looking at each of the types of problem found in the study, we observed all the ordinal problems were simpler with some having lines of irrelevant information.

When looking at the percentage answering correctly, CP proved to be the most challenging. While there may be a contribution from the type, the questions having much higher ambiguity in CP is more likely the cause.

The candidates were told that some question would be unsolvable.



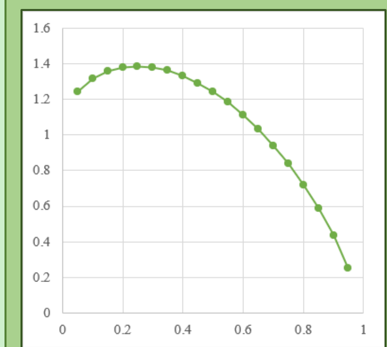
There were also different percentages correct depending on whether the question was a distractor (S1D,S2D) or a solvable question (S1S,S2S).

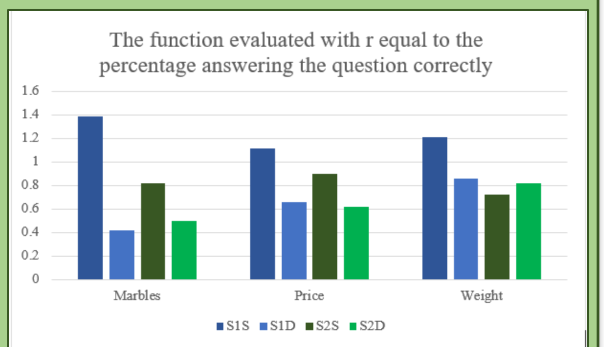
Entropy and Word Problems

By using a tree to map out the possible routes of interpretation, there are many points at which the entropy can be evaluated. After each new path, at the end of a completed path, the sum of each new paths. The entropy calculation we decided on was for the total tree and used the probability of reaching each end point to calculate. In the equation below r is the product of the probability p1 and p2. Where p1 is the first branch and p2 the second.

$$-r\ln r\ -\ (1-r)\ln(1-r) + (1-r)\ln 3$$

In the first modelling attempt a similar formula was found which did not include the last term, (1-r)ln3, the equation did not calculate the entropy of the system but the surprise of it. The creation of the r term was a result of trying to use a single probability value, however due to the laws of probability, the product of two probabilities cannot be greater than the lowest of the two. (0.5*P =< 0.5). The plot below shows the function for a range of values and the chart shows it evaluated with the r values from the cardinal problems.





Conclusion

The end result of the project was a formula that works for the cardinal example. With some tweaking and further research, the model can be expanded to work on other questions with ambiguity.

The next step forward is to continue working on the model and find another study where interpretation of text is measured. Applying the model to other studies data would allow us to find other issues with the model and correct them.

Reference

****https://link.springer.com/article/10.3758/s13423-019-01628-3**

Student: Margaret-Ann Withington

Supervisor: Dr Fabien Paillusson, Dr Matthew Booth & Dr Manuela Mura

UNIVERSITY OF LINCOLN
Lincoln Academy of Learning and Teaching

Undergraduate Research Opportunities Scheme

DISCOVER_
LNCN.AC/UROS
@UOL_LALT
#UROS2021