

Correlation Analysis on Generalization Capacities of Different Compositional Problems

Group Members: Yoobin Cheong, Yeong Koh, Yoon Tae Park

What are Compositional Problems?

Train smaller parts
that the data can be decomposed into

[The girl], [The cat], [The boy]
[The cat **loves** the girl]
[The **hedgehog** see the cat]



Generalize data
by combining known, smaller parts
[The boy **loves** the **hedgehog**]

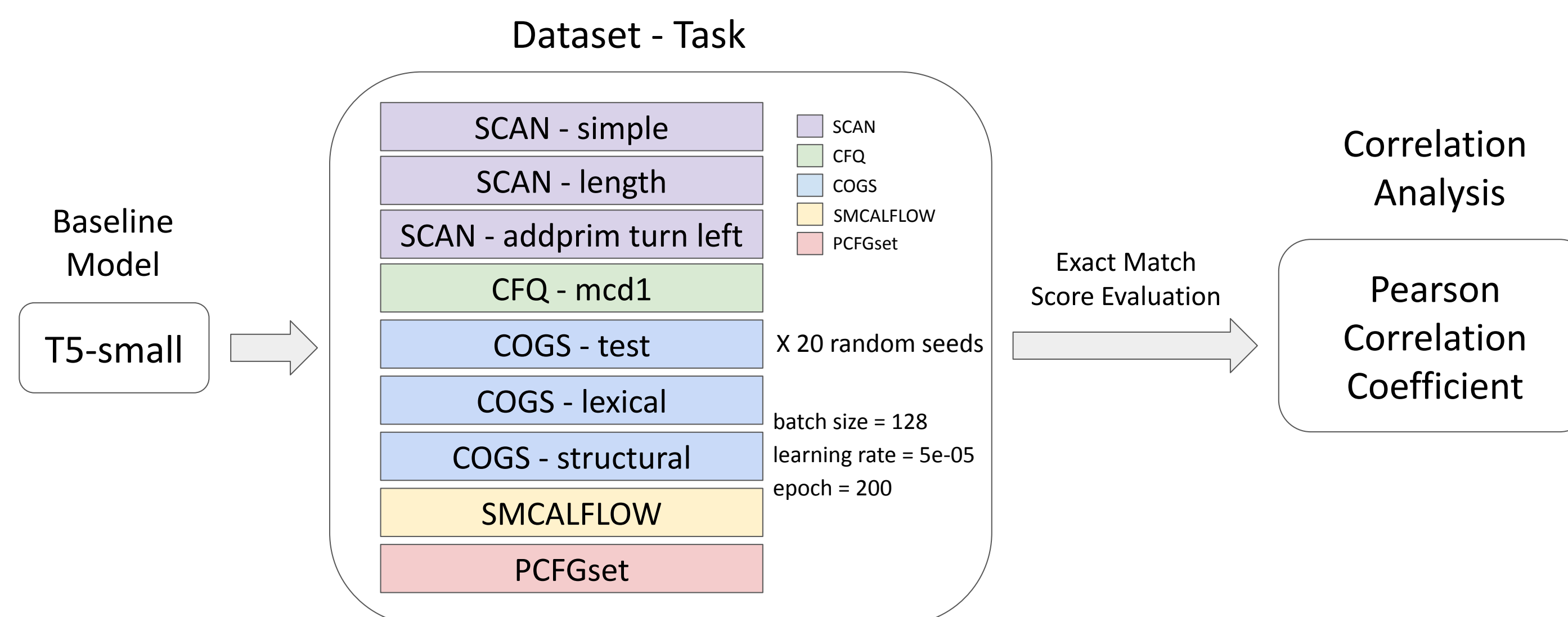
Why this matters? Compositional generalization capacity is often presented as a solution to generalize outside of the training data, but it is also considered that neural networks struggle to achieve

Main Goal: Experiment on multiple tasks to evaluate performance given different weight initializations, and conduct correlation analysis between different tasks to see that if there are any evidence of compositional generalization capacity

Q: Do **Different** Compositional Problems Rely on **Similar** Capacities?

A: For some tasks, yes. In general, need to investigate more.

Experimental Design



Step 1 Use different weight instantiations of the non-pretrained t5 model and **calculate exact match score** for **20 different** random seeds for given task

Step 2 **Extend to other datasets** that we defined and also calculate exact match score

Step 3 Calculate **Pearson Correlation Coefficient** for all different combination of task pairs
(X: selected task 1, Y: selected task 2, X != Y)

$$\rho_{X,Y} = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

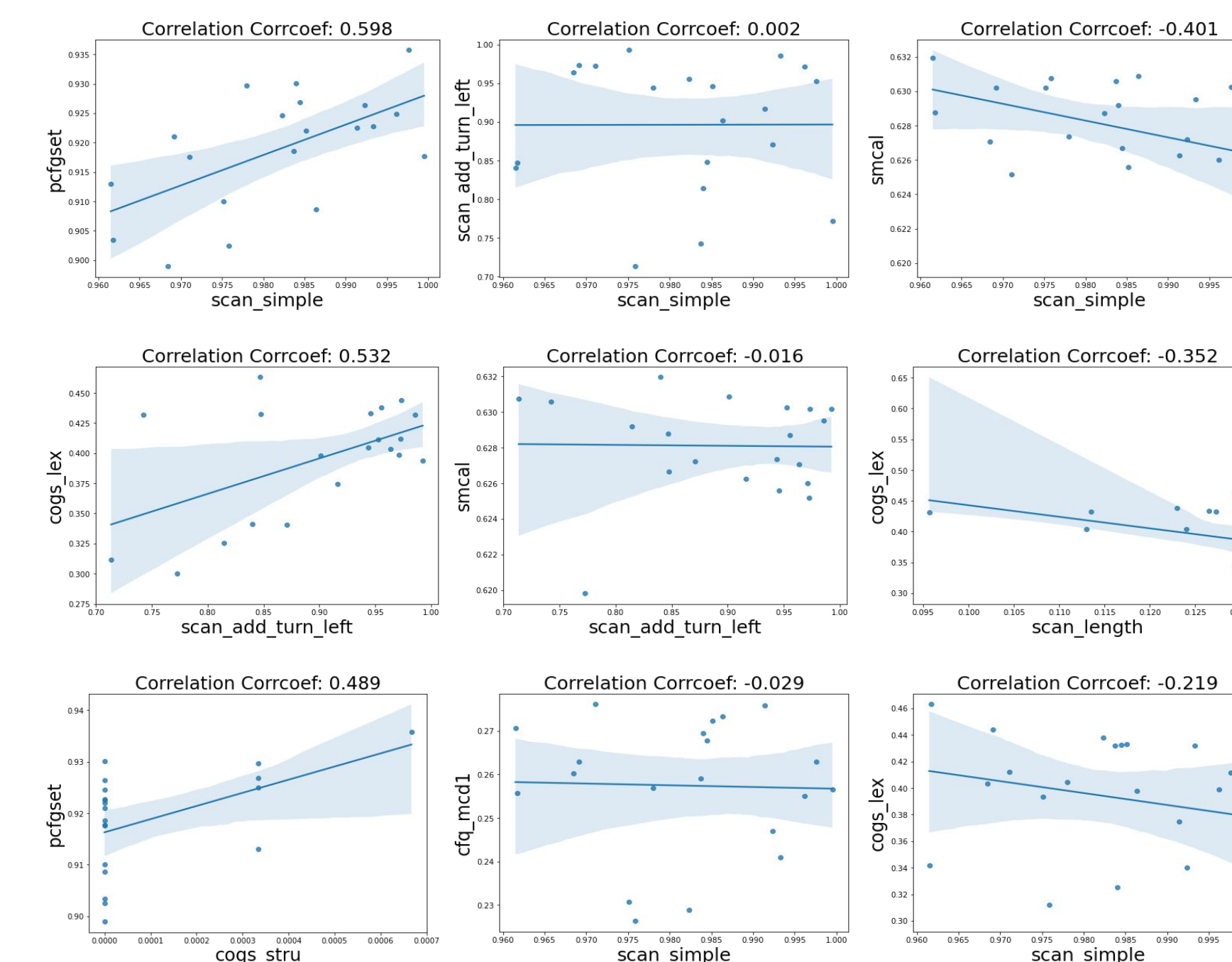
Results

We've conducted 36 experiments which consist of 9 different tasks with various datasets. The **mean correlation** across all pairs of tasks was **0.186**. Below are the main findings.

[Datasets with Relatively High Correlations]

Dataset 1	Dataset 2	Correlation Coefficient
SCAN-simple	PCFGset	0.598
SCAN-addprim turn left	COGS-lexical	0.532

[Top3 Correlations on Positive, Neutral, and Negative]



Conclusion

We've found that there were some highly correlated tasks, evidencing that those tasks would share similar compositional generalization capacities. We believe that further experimenting on those tasks will help to solve compositional problems that current SOTA transformer models struggle with.