Experiment list

I have 174 sentences and 3467 tokens annotated, filtered and checked.

Because of how supar does batching, the number of tokens is actually what's meaningful for splitting up the data into batches. However, note that the batching will not split apart a sentence so if the batch size is set to 8, the result will be one sentence per batch because none of my sentences are less than 8 words in length.

Does augmentation with rule generated data help at all in this scenario?

20 runs total

- Baseline 5 fold cross validation using only the 175 sentences.
- Balanced inclusion of 175, 350, 525, 700 augmented sentences with full trees.

Is curriculum learning useful in a scenario where both quality and difficulty are considered?

5-fold cross validation between two difficulty functions and three curriculum: 30 total runs.

Difficulty functions:

- Quality is modeled as an offset where some value p is added to the difficulty of the sentence. (probably a negative offset because the higher quality sentences should probably come later).
- Quality is modeled as an orthogonal aspect where the standard bucketing is used for length/difficulty clustering but the buckets also encode information about quality

Curriculum:

- homogeneous interleaved batches
- heterogeneous linear introduction
- homogeneous series (one then the other)

Is it better to use rule generated augmented data or augmented data derived from the output of some parser or a group of parsers?