Annotations

Eneko Agirre

- A random sample of 64 pairs from MSR video and MSR paraphrase (mixed up)

- A random sample of 51 pairs from SMT

- 23 hand-picked sample of pairs from MSR video

  - some pairs where one participant was concerned about the turkers annotations

For discussion

- Is the likert scale with 6 values about right?

- Are you happy with the instructions?

- Dataset by dataset:

    - Are there examples which do not fit well in the task?

    - Could you characterize phenomena you took into account?

- Is there a trend across different datasets?