Basic Manual Evaluation Decisions

What to Show to the annotators when assessing the candidate?
- REF-based … only the (human) reference
- SRC-based … only the source
- SRC&REF-based … both

Context to Consider:
- Sentence-level … sentences in random order
- Document-level … obtain single score per document
- Document-aware … show whole documents, scores per sentence

What to Ask from annotators (scoring technique):
- Some relative score over several candidates?
- Some absolute score for a single output?
- A more complicated question?