We propose a targeted textual entailment task designed to train and evaluate parsers. Recent approaches on cross-framework parser evaluation employ framework-independent representations such as GR and SD schemes. However, there is still arbitrariness in the definition of such a scheme and the conversion is problematic. Our approach takes this idea one step further. Correct parse decisions are captured by natural language sentences called textual entailments. Participants make a yes/no choice on a given entailment. It will be possible to automatically decide which entailments are implied based on the parser output only, i.e. there will be no need for lexical semantics, anaphora resolution etc.

- Final-hour trading accelerated to 108.1 million shares, a record for the Big Board.
  - 108.1 million shares was a record. – YES
  - Final-hour trading accelerated a record. – NO

The proposed task is desirable for several reasons. First, textual entailments focus on the semantically meaningful parser decisions. Trivial differences are abstracted away, which should result in a more accurate assessment of parser performance on real-word applications. Second, no formal training is required. Annotation will be easier and annotation errors will have a less detrimental effect on evaluation accuracy. Finally, entailments will be non-trivial since they will be collected by considering the differences between the outputs of different state-of-the-art parsers.

The participants will be provided with training and test sets of entailments and they will be evaluated using the standard tools and methodology of the RTE challenges. External training data may be used in addition to the provided training data. Modules that derive all textual entailments given a parser output in Penn Treebank format for phrase structure parsing and in CoNLL format for dependency parsing will be made publicly available. Therefore majority of parser developers can participate in the task with little effort and are indeed invited to do so.