

# Tools Used in Creating the Concretely Annotated Corpora

Documents contained in this dataset have the following language annotations: 1. Segmented sentences and Penn Treebank-style tokenized words 2. Treebank-style constituent parse trees 3. Four different syntactic dependency trees 4. Named entities 5. Part of speech tags 6. Lemmas 6. In-document entity coreference chains, and 7. Three different frame semantic parses.

concrete-stanford, a wrapped version of Stanford's CoreNLP (version 3.5.2), segmented the sentences and words (1), produced the constituent parse (2), three of the dependency parses (3), and the named entities, part of speech tags, lemmas, and an entity coreference chain (4-7).

concrete-semafor, a wrapped version of the SEMAFOR frame semantic parser (version 2.1), produced the fourth dependency parse (3), Entity mentions (6) and one semantic parse (7).

fnparse, a semantic role labeling system, produced the PropBank and one of the FrameNet semantic parses (7).