A large annotated corpus of entailments and contradictions
Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning
Stanford University

Corpus methods have the potential to offer a range of fast reproducible experimental
designs to researchers in semantics. However, these methods have been largely impractical
due to a lack of large corpora with any kind of semantic grounding. While there are serious
obstacles to collecting such data, we argue that focusing on inferences of entailment and
contradiction as a form of grounding mitigates these, and we present a large new corpus of
utterance pairs labeled for the entailment and contradiction, the Stanford Natural Language
Inference corpus (introduced earlier this year in Bowman et al., 2015).

An ideal corpus for model-theoretic semantics would pair utterances in context with
either a truth value or, even better, a set of truth conditions expressed in a logical form of
some kind. Doing this requires specifying a representational system to describe the contexts
of utterances, and in the latter case, the contents of those utterances as well. How best to
represent this kind of information, though, is very much an open problem, and in fact it is
the very problem that such a corpus would be meant to help solve.

Following work in natural logic (Moss, 2009) and focusing on grounding the meanings
of utterances in inferences that they licence, rather than in their truth conditions, makes
it possible to collect a representation-agnostic corpus. In this framing, natural language
is treated as its own representation language, and the semantic annotation for a sentence
consists simply of examples of sentences with respect to which it is entailing, contradictory,
or semantically neutral. High quality corpora of this kind exist (Dagan et al., 2006, Marelli
et al., 2014), but at a few thousand examples they have so far been too small for corpus
methods to offer much insight. Our new corpus of 570k examples aims to remedy this.

In our data collection effort, annotators recruited through Amazon Mechanical Turk
were presented with short scene descriptions (either sentences or complex NPs), and asked
to redescribe each scene first faithfully (yielding a pair with the label entailment), then
deliberately unfaithfully (contradiction), and then with the introduction of uncertain details
(neutral). We collected 570k such pairs of utterances, and had additional annotators relabel
57k of these, including those shown in Table 1. We find the resulting pairs reliably evince
valid semantic relationships between sentences with diverse surface forms, and we hope that
they can serve as a valuable resource in exploratory and experimental research on meaning.
They are available for download at: nlp.stanford.edu/projects/snli/

Table 1: Randomly chosen examples from our corpus, shown with both labels from individual
annotators and consensus labels.
References


