Causal relations represent a significant fraction of the intermediate-level semantics of a text. To assist with downstream semantic processing, I am defining an annotation scheme for the causal relations that are expressed by a text, including such attributes as the degree/polarity and the type of causation. I am working with annotators to build a small annotated corpus, which will serve to validate the representation. I am also building a machine learning system that will be trained and tested on this corpus. The system will use a variety of features, particularly TRegex patterns on the dependency parse of the sentence, to identify and classify unseen instances of causal relations.
Argentina played another great game last night against Brazil — the last game in Group C. In the first half, Brazil maintained the majority of the possession, but got few clear looks on goal. With under 5 minutes to half-time, a long ball was sent towards the Brazilian center back, who misplayed the ball. As the ball skipped behind him, the Argentinian center forward was able to gather it for a 1v1 with the keeper. The result was a 1-0 lead for the Argentina. In the second half, Brazil struggled to keep possession of the ball. The Argentinian midfielders were able to stop any clear looks, preserving the shutout and advancing their team to the knockout stages.
This was the decisive game in qualifying rounds, a tense encounter between Argentina and Brazil. For the first half of the game, Brazil seemed to be running the show. But halfway through, the Brazilians made a mistake, which an Argentinian player exploited to circumvent Brazil’s defenses and pull into the lead. The Brazilians never recovered: the Argentinian team then kept up their own defenses for the rest of the game, and ultimately won the game!

It would take you about 20 years to read the articles we publish online every day. The only way we’ll be able to harness this information overload is if our computers can help us extract the key information – but even state-of-the-art software can interpret only the simplest English statements. As a first step toward more sophisticated interpretation, I’m building a system to extract one of the most common and useful types of information: statements about cause and effect (e.g., “smoking leads to cancer”). To teach the system what such expressions look like, I am first building a collection of documents in which humans have manually flagged and analyzed cause-and-effect language. From these documents, the system will learn to extract different kinds of cause-and-effect language, ultimately allowing us to see at a glance the key arguments being made.