

Introduction to the Stanford POS Tagger

Sean Stewart - 19 October 2012

Check back for an update regarding how we used Stanford's tagging program written in Java!

Stanford Part-of-Speech Tagger

Sean Stewart - 18 October 2012

For our project, Beth and I are going to use the part of speech tagger made by Stanford, called the Stanford Part-of-Speech tagger. For our research, we wish to find all of the verbs used by Jules Verne in his work "Vingt-mille lieues sous les mers" *Twenty-thousand Leagues Under the Sea*. In order to do so, we could have read the entire text and marked each verb's tense and aspect. Instead, we have acquired a pre-marked text and will send this marked-text through the Stanford POS tagger, which will spit out, for example a string such as: `A_DT passenger_NN plane_NN has_VBZ crashed_VBN shortly_RB after_IN take-off_NN`.

We will then analyse each verb (denoted as `_V` in French) to see which form (literary or non-literary) it has been written in. Ultimately it is our goal to analyse the text as a whole in order to see if any of Verne's characters use the literary past tense during quotations or whether the literary past is solely used in Verne's own narration of past events.