Mohammad vs Nick

Mohammad and Nick have an excellent agreement rate and an excellent Cohen’s Kappa as well. Their disagreements come on the easily controversial tags. For example, on PROPN they have no agreement; however this appears to be due to how Nick tagged the organization names in the text, opting to have every word in the organization name be a proper noun. Mohammad took the alternative approach which is to tag every word as its proper part of speech independent of whether it is part of a name. Their other disagreements were on similar types of issues where the confusion is easily explained. They had minor disagreement on ADP and DET, though in French this is a common confusion because of terms like “de” which can in some contexts mean “from” and in others be used to express that one has “some” of something. Additionally there was one disagreement where Nick said that something was a CONJ but Mohammad put SCONJ; both recognized the conjunction yet disagreed on whether it was subordinate. Overall, their agreement is excellent by any of the proposed metrics for assessing a Cohen Kappa.

Nick vs Arif

Nick and Arif have fairly good agreement rate. The majority of their disagreement comes from the fact that Arif omitted tags where he was completely unable to decipher the part of speech and he put in X for blank words where Nick simply left these blank. On the majority of tags they do have good agreement with confusion coming on the objectively confusing words. For example, the word “le/la/l’” preceding verbs was a point of contention with Arif labeling it a DET and Nick labeling it a PRON. This makes sense since in those situations “le/la/l’” takes the place of the object, but normally “le/la/l’” is a determiner. Another agreement issue, which is common to many of the annotation differences, is what words are PROPNs. With organization names and names containing numbers there is much room for disagreement about which tag is correct. Other disagreements were minor such as AUX vs VERB and CONJ vs SCONJ. Some amount of error can certainly be attributed to the fact that neither annotator is a native speaker. Overall, by Landis and Koch’s guidelines the pair had substantial agreement and by Fleiss’ standards the pair had almost excellent agreement.