**The Conlleval metrics is better than the Python Script for the NER task.**

Explanation :

The Conlleval task divides the text into syntactically related non-overlapping group of words or text-chunks. Text-chunks helps in retaining a group of syntactically related words as a phrase. These phrases are non-overlapping which means that one word belongs to only one chunk and hence it is easier for the NER task which identifies every group as an entity.

For eg : He<NP> reckons<VP> [the current account deficit]<VP>

And the corresponding numbers of accuracy and FB1 score give more accurate results than the python script based approach which only takes into account the individual tokens which may be overlapping. Thus rendering it to be less accurate.