In criminal proceedings, it is often difficult to assess the sincerity of testimonies. Methods helping this assessment would provide useful support to criminal investigations.

We developed stylometric techniques to classify statements as false or true. Our data are not 'mock lies' produced in laboratory conditions, but genuine Court data (the first corpus of this type in Italy).

Our corpus collects criminal proceedings of cases of calumny and false testimony in which the Court record contains:

1) verbatim transcriptions of testimonies collected during the hearings;

2) a final Court judgment of guilty that describes the events and identifies the untruthful statements issued by the defendant.

Thanks to this information, we were able to annotate the utterances in the testimonies as false, true, or uncertain with a much greater degree of confidence than ever before.

This corpus was used to train models able to distinguish true from false utterances using supervised machine learning algorithms. At first sight the task might seem almost impossible, considering, e.g., the great number of very short statements; but as already seen in other applications of stilometryc techniques, the accuracy in the classification task was about 70%, clearly higher than chance level.