A Setup configuration used to train our sequence labeling methods

**Conditional Random Fields**  We use the default configuration provided together with the sklearn-crfsuite library.

**MultiLayer Perceptron**  Both the discrete and distributed perceptrons are implemented in keras.

- **Training hyperparameters**  The model is trained up to 30 epochs, with early stopping (patience=4). We use Stochastic Gradient Descent (SGD) to optimize the objective function. The initial learning rate is set to 0.1.

- **Layer and embedding sizes**  The dimension of the hidden layer is set to 100. For the perceptron fed with embeddings, we use 100 and 20 dimensions to represent a word and its PoS tag, respectively.

**Bidirectional Long Short-Term Memory**  We relied on the NCRFpp framework (Yang and Zhang, 2018).

- **Training hyperparameters**  We use mini-batching (the batch size during training is set to 8). As optimizer, we use SGD, setting the initial learning rate to 0.2, momentum to 0.9 and a linear decay of 0.05. We train the model up to 100 epochs and keep the best performing model in the development set.

- **Layer and embedding sizes**  We use 100, 30 and 20 dimensions to represent a word, a postag and a character embedding. The output hidden layer from the character embeddings layer is set to 50. The left-to-right and right-to-left LSTMs generate each a hidden vector of size 400.