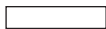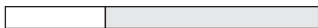Sequence of neighbouring words
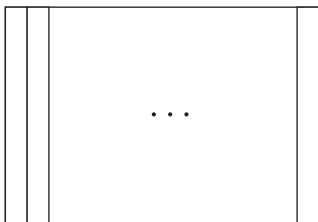$(2\ n\_words\_left\_right + 1)$

BERT encoding

Encoded and padded sequence (128)

BERT embedding layers

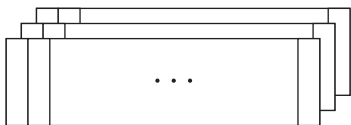BERT embeddings (768 x 128)

$\cdots$

3 convolutional layers (ker. size = 1, 2, 3)

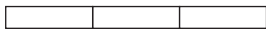($conv\_filters$ x 127, $conv\_filters$ x 126, $conv\_filters$ x 125)

$\cdots$

Max pooling

($conv\_filters$ x 1, $conv\_filters$ x 1, $conv\_filters$ x 1)

Concatenation

(3 $conv\_filters$)

Dense layer

($dense\_units$)

Dropout

($dense\_units$)

Dense layer

(5)