Appendix A4: Details of weight computation in *SyLF-LPBridge* classifier

Quadruples on the shortest path in the bridge representation of a sentence form the features. Words, links and thereby the quadruples in a bridge need to be weighed differently based on their importance in connectivity extraction. Consider the sentence, *"All nerves including those innervating the BR1, project to the BR2 and are somatotopically organized."*. The bridge for this sentence is, *[['Os', 'innerv', 'BR1', 1], ['Mg', 'those', 'innerv'], ['Op', 'includ', 'those', 1], ['MX*p', 'n', 'includ', 1], ['Sp', 'n', 'and', 2], ['VJlpi', 'project', 'and', 2], ['MVp', 'project', 'to', 2], ['Js', 'to', 'BR2', 2]]* , where the noun *'nerves'* is denoted as *'n'*.
The weight of a quadruple *Q=<Link(l), Left word(lw), Right word(rw), Context(c)>* is based on the weights of the words and links making it up. Weight of a word or link e is based on the importance of e in the particular context c and is calculated using Eq. 1. The discrimination ability of e with respect to classifying the connection between the brain regions is used to calculate the importance of e. This can be looked upon as a kind of entropy measuring the discernibility. Higher the discernibility, lower the entropy. Frequency of occurrence is used to scale the entropy to arrive upon the weight of the link/word e. Thus, weights of all links and words in the training data are computed for each of the contexts.

$$\begin{aligned} weight(e) &= log\left[Freq(e) * \frac{1}{Ent(e)}\right] \\ Freq(e) &= log(frequency(e)) \\ Ent(e) &= -PlogP - NlogN \end{aligned} \tag{1}$$

where P and N are the ratio of positive sentences and negative sentences with e respectively, computed over context c in the training corpus and *frequency(e)* is the number of times e has occurred in the corresponding context c.
Weight of the quadruple $Q =< l, lw, rw, c >$ is calculated as in Eq.2.

$$weight(Q) = \frac{weight(l) + weight(lw) + weight(rw) + weight(c)}{4} \tag{2}$$

where weights for l, lw and rw are calculated using Eq.1 and the choice of weights for c is explained in Appendix A3.
The final Bridge representation for the sentence will be,
[['Os', 'innerv', 'BR1', 1], ***0.37****], [['Mg', 'those', 'innerv', 1],* ***0.44****], [['Op', 'includ', 'those', 1],* ***0.59****], [['MX*p', 'n', 'includ', 1],* ***0.25****], [['Sp', 'n', 'and', 2],* ***0.43****], [['VJlpi', 'project', 'and', 2],* ***0.66****], [['MVp', 'project', 'to', 2],* ***0.68****], [['Js', 'to', 'BR2', 2],* ***0.44****]].*
It can be noticed that informative quadruples like *['VJlpi', 'project', 'and', 2]* and *['MVp', 'project', 'to', 2]* which are important connectivity indicators have higher weights compared to quadruples *['MX*p', 'n', 'includ', 1]* and *['Os', 'innerv', 'BR1', 1].*