

Considerando o conceito de *near misses*, algumas soluções foram propostas. Proposta por [?], P_k atribui valores parciais a *near misses*, ou seja, limites sempre receberão um peso proporcional à sua proximidade, desde que dentro de um janela de tamanho k . Para isso, esse método move uma janela de tamanho k ao longo do texto. A cada passo verifica, na referência e na hipótese, se as extremidades (a primeira e última sentença) da janela estão ou não dentro do mesmo segmento, então, penaliza o algoritmo caso este não concorde com a referência. Ou seja, dado duas palavras de distância k , o algoritmo é penalizado caso não concorde com a segmentação de referência se as palavras estão ou não no mesmo segmento.

Dadas uma segmentação de referência ref e uma segmentação automática hyp , ambas com N sentenças, P_k é computada como:

$$P_k(ref, hyp) = \frac{1}{N - k} \sum_{i=1}^{N-k} (\delta_{ref}(i, i + k) \oplus \delta_{hyp}(i, i + k)) \quad (1)$$

onde $\delta_S(i, j)$ é a função indicadora que retorna 1 se as sentenças i e j estão no mesmo segmento e 0 caso contrário, \oplus é o operador **XNOR** (ou exclusivo) que retorna 1 se ambos os argumentos forem diferentes. O valor de k é calculado como a metade da média dos comprimentos dos segmentos reais. Como resultado, é retornado a dissimilaridade entre as segmentação calculada pela contagem de discrepâncias dividida pela quantidade de segmentações analisadas. Essa medida pode ser interpretada como a probabilidade de duas sentenças extraídas aleatoriamente pertencerem ao mesmo segmento.