Both the *SinSRL: Projector* tool and the *SinSRL: Direct Annotator* tool are evaluated separately. Precision,Recall and F1 score is used to evaluate the semantically tagged sentences. especially, the evaluation is done per entity (per span) based not per token. The equations of calculating F1-Score are as follows.

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

Figure 24 - Definition of Precision

Precision is the number of correct results divided by the number of all returned results.

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

Figure 25 - Definition of Recall

Recall is the number of correct results divided by the number of results that should have been returned.

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Figure 26 - Definition of F1 Value

Other than the F1 score, a new marking scheme was used to evaluate tools. Marking scheme for the evaluation is as follows.

Marks for each sentence = 100

I.e. පලස්තීන සිවිල් වැසියන් හමාස් වලින් ආරක්ෂා කළ යුතුය . ⇒ marks = 100

Marks for each detected predicate = 100/n ; n - no. of predicates

I.e [පලස්තීන ARG1 , සිවිල් ARG1 , වැසියන් ARG1 , හමාස් ARG2 , වලින් ARG2 , ආරක්ෂා protect.01 , කළ protect.01 , යුතුය O , . O] , [පලස්තීන ARG0 , සිවිල් ARG0 , වැසියන් ARG0 , හමාස් ARG1 , වලින් ARG1 , ආරක්ෂා ARG1 , කළ ARG1 , යුතුය need.01 , . O] ⇒ marks = 100/2

Marks for a single sequence are divided as follows.

[පලස්තීන ARG1 , සිවිල් ARG1 , වැසියන් ARG1 , හමාස් ARG2 , වලින් ARG2 , ආරක්ෂා protect.01 , කළ protect.01 , යුතුය O , . O] ⇒ marks = m (Here m = 100/2 and 5 marks are deducted for extra ones)

Marks for predicate tags = m/2 Marks for SRL tags = m/2

No. of predicate tags = 2 (Here ආරක්ෂා protect.01 , කළ protect.01)

No. of SRL tags = no.of all the tags - no. of predicate tags

Marks of a single predicate tag are divided into 90% for the predicate verb and 10% for the sense.

Marks for a single SRL tag = (m/2) / No. of SRL tags

Overall accuracy is obtained by calculating average marks.