Table 1: Confusion matrix using Gilbert Annotation Schemes

Annotator 1 → Annotator 2 ↓	HATE	NOHATE
HATE	11	0
NOHATE	6	27

$po \ = \ \frac{11+27}{44} \ = \ 0.86$

$pe:$

$Annotator\ 1\ said\ 17\ to\ "HATE"\ and\ 27\ to\ "NOHATE".\ Thus\ Annotator\ 1\ said\ "NOHATE"\ 60\%\ of\ the\ time.$

$Annotator\ 2\ said\ 11\ to\ "HATE"\ and\ 33\ to\ "NOHATE".\ Thus\ Annotator\ 2\ said\ "NOHATE"\ 75\%\ of\ the\ time.$

$pHATE \ = \ \frac{11}{44} \cdot \frac{17}{44} \ = \ 0.09$

$pNOHATE \ = \ \frac{33}{44} \cdot \frac{27}{44} \ = \ 0.46$

$pe \ = \ pHATE \ + \ pNOHATE \ = \ 0.55$

$\kappa \ = \ \frac{po - pe}{1 - pe} \ = \ 0.68$

Table 2: Confusion matrix using Kumar Annotation Schemes

Annotator 1 → Annotator 2 ↓	CAG	NAG	OAG
CAG	7	1	1
NAG	0	14	3
OAG	4	2	12

$po \ = \ \frac{7 + 14 + 12}{44} \ = \ 0.75$

$pe:$

$Annotator\ 1\ said\ 11\ to\ "CAG",\ 17\ to\ "NAG"\ and\ 16\ to\ "OAG".\ Thus\ Annotator\ 1\ said\ "NAG"\ 39\%\ of\ the\ time.$

$Annotator\ 2\ said\ 9\ to\ "CAG",\ 17\ to\ "NAG"\ \ and\ 18\ to\ "OAG".\ Thus\ Annotator\ 2\ said\ "OAG"\ 41\%\ of\ the\ time.$

$pCAG \ = \ \frac{9}{44} \cdot \frac{17}{44} \ = \ 0.08$

$pNAG \ = \ \frac{17}{44} \cdot \frac{17}{44} \ = \ 0.15$

$pOAG \ = \ \frac{18}{44} \cdot \frac{16}{44} \ = \ 0.15$

$pe \ = \ pCAG \ + \ pNAG \ + pOAG \ = \ 0.38$

$\kappa \ = \ \frac{po - pe}{1 - pe} \ = \ 0.60$

Table 3.1: Confusion matrix using Zamp Annotation Schemes - top level annotation

Annotator 1 → Annotator 2 ↓	OFF	NON
OFF	14	7
NON	4	19

$po = \frac{14+19}{44} = 0.75$

$pe:$

$Annotator\ 1\ said\ 18\ to\ "OFF"\ and\ 26\ to\ "NON".\ Thus\ Annotator\ 1\ said\ "NON"\ 59\%\ of\ the\ time.$

$Annotator\ 2\ said\ 21\ to\ "OFF"\ and\ 23\ to\ "NON".\ Thus\ Annotator\ 2\ said\ "NON"\ 52\%\ of\ the\ time.$

$pOFF = \frac{21}{44} \cdot \frac{18}{44} = 0.19$

$pNON = \frac{23}{44} \cdot \frac{26}{44} = 0.30$

$pe = pOFF + pNON = 0.49$

$\kappa = \frac{po - pe}{1 - pe} = 0.50$

Table 3.2: Confusion matrix using Zamp Annotation Schemes - target annotation

Annotator 1 → Annotator 2 ↓	TARG	NOTARG
TARG	11	1
NOTARG	1	1

$po = \frac{11+1}{14} = 0.85$

$pe:$

$Annotator\ 1\ said\ 12\ to\ "TARG"\ and\ 2\ to\ "NOTARG".\ Thus\ Annotator\ 1\ said\ "TARG"\ 85\%\ of\ the\ time.$

$Annotator\ 2\ said\ 12\ to\ "TARG"\ and\ 2\ to\ "NOTARG".\ Thus\ Annotator\ 2\ said\ "TARG"\ 85\%\ of\ the\ time.$

$pTARG = \frac{12}{14} \cdot \frac{12}{14} = 0.73$

$pNOTARG = \frac{2}{14} \cdot \frac{2}{14} = 0.020$

$pe = pTARG + pNOTARG = 0.75$

$\kappa = \frac{po - pe}{1 - pe} = 0.4$

Table 3.3: Confusion matrix using Zamp Annotation Schemes - target group annotation

Annotator 1 → Annotator 2 ↓	G	I	O
G	8	1	0
I	0	2	1
O	0	0	0

$$po\ =\ \frac{8+2+0}{12}\ =\ 0.83$$

$Annotator\ 1\ said\ 8\ to\ "G", 3\ to\ "I"\ and\ 1\ to\ "O".\ Thus\ Annotator\ 1\ said\ "G"\ 67\%\ of\ the\ time.$

$Annotator\ 2\ said\ 9\ to\ "G", 3\ to\ "I"\ and\ 0\ to\ "O".\ Thus\ Annotator\ 2\ said\ "G"\ 75\%\ of\ the\ time.$

$$pG\ =\ \frac{9}{12}\cdot\frac{8}{12}\ =\ 0.5$$

$$pI\ =\ \frac{3}{12}\cdot\frac{3}{12}\ =\ 0.06$$

$$pO\ =\ \frac{0}{12}\cdot\frac{1}{12}\ =\ 0$$

$$pe\ =\ pG\ +\ pI\ +\ pO\ =\ 0.56$$

$$\kappa\ =\ \frac{po-pe}{1-pe}\ =\ 0.61$$

From these results, it can be seen that all classifications are **reliable** as the inter-rater acceptance test passes **75%**. However, stronger inter-rater reliability might still be required, though. Because it considers the likelihood that the agreement may occur by chance, is typically thought to be a more reliable measurement than the standard percentage agreement calculation.

When the calculated Kappa values are examined, it is seen that the “HATE-NOHATE” result is “Substantial Agreement”, “CAG-NAG-OAG” result is “Moderate Agreement”, “OFF-NON” result is “Moderate Agreement”, “TARG-NOTARG” result is “Fair Agreement” and finally “G-I-O” is “Substantial Agreement”. As can be seen from these results, “Near Perfect Agreement” and “Perfect Agreement” are not included in the reliability measurement results among the annotator agreements. The most reliable results are seen in the “HATE-NOHATE” and “G-I-O” annotations.