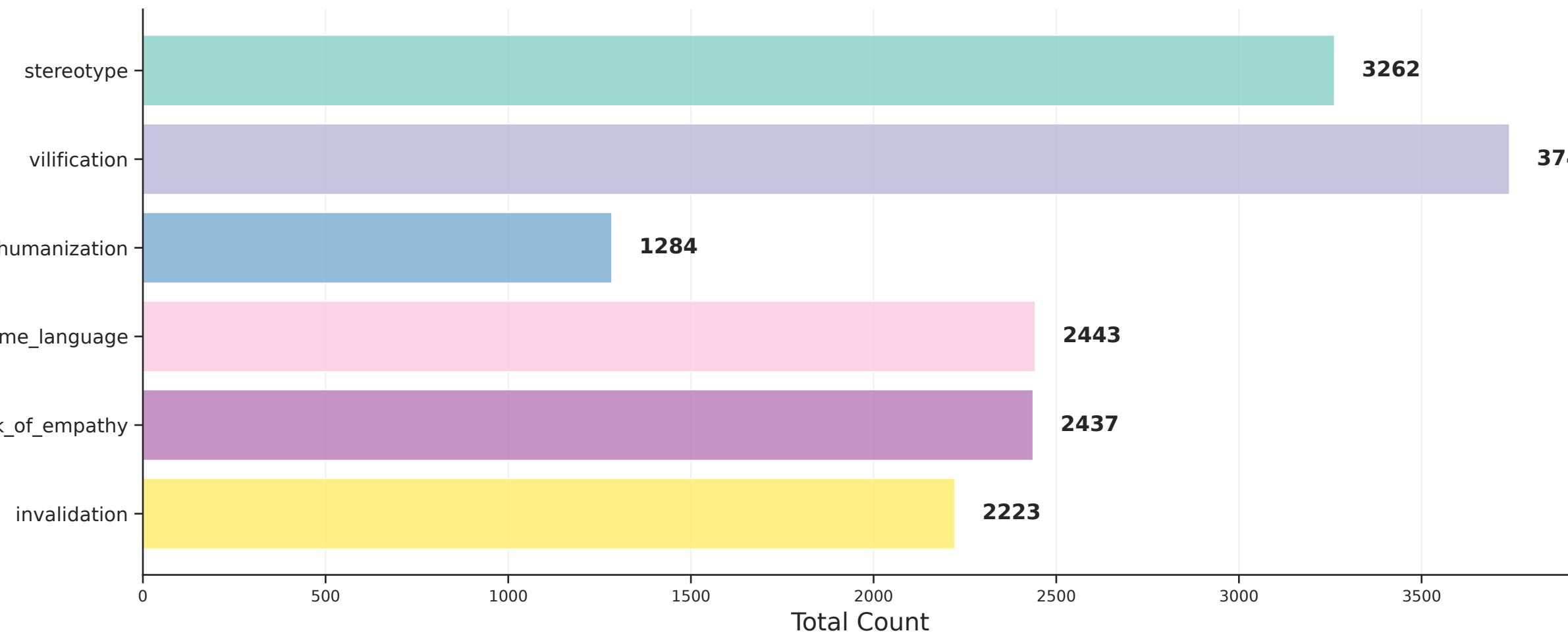
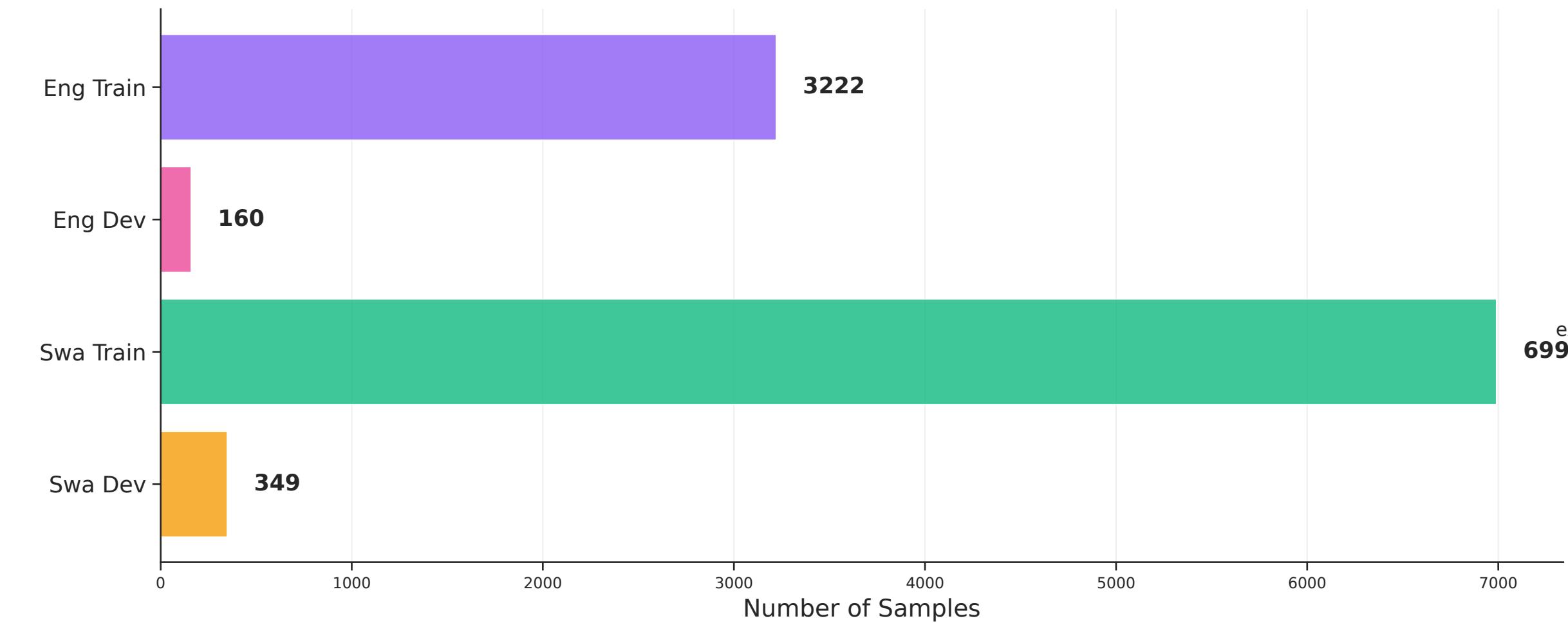


Dataset Sizes

Subtask 3 - Data Distribution Analysis (Multi-Label)

Overall Multi-Label Distribution

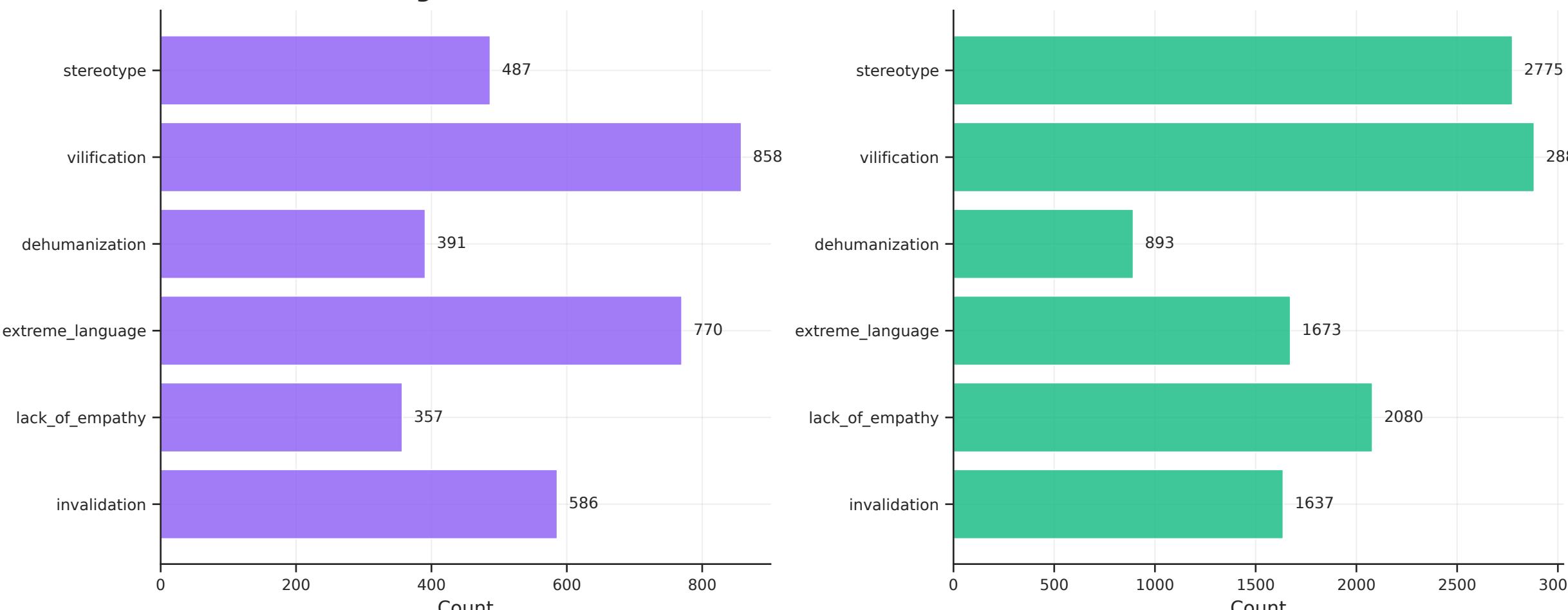


English Train

Swahili Train

Label Co-occurrence

Label Positive Rates



Count

0

200

400

600

800

Swahili Train

Count

0

500

1000

1500

2000

2500

Label Co-occurrence

Count

0

1000

2000

3000

4000

5000

6000

7000

8000

9000

10000

11000

12000

13000

14000

15000

16000

17000

18000

19000

20000

21000

22000

23000

24000

25000

26000

Label Positive Rates

Positive Rate (%)

Count

0

500

1000

1500

2000

2500

3000

3500

4000

4500

5000

5500

6000

6500

7000

7500

8000

8500

9000

9500

10000

10500

11000

11500

12000

12500

13000

13500

14000

14500

15000

15500

16000

16500

17000

17500

18000

18500

19000

19500

20000

English: Train vs Dev

Count

Labels

Train

Dev

stereotype

vilification

dehumanization

extreme_language

lack_of_empathy

invalidation

Count

Labels

Train

Dev

Swahili: Train vs Dev

Count

Labels

Train

Dev

stereotype

vilification

dehumanization

extreme_language

lack_of_empathy

invalidation