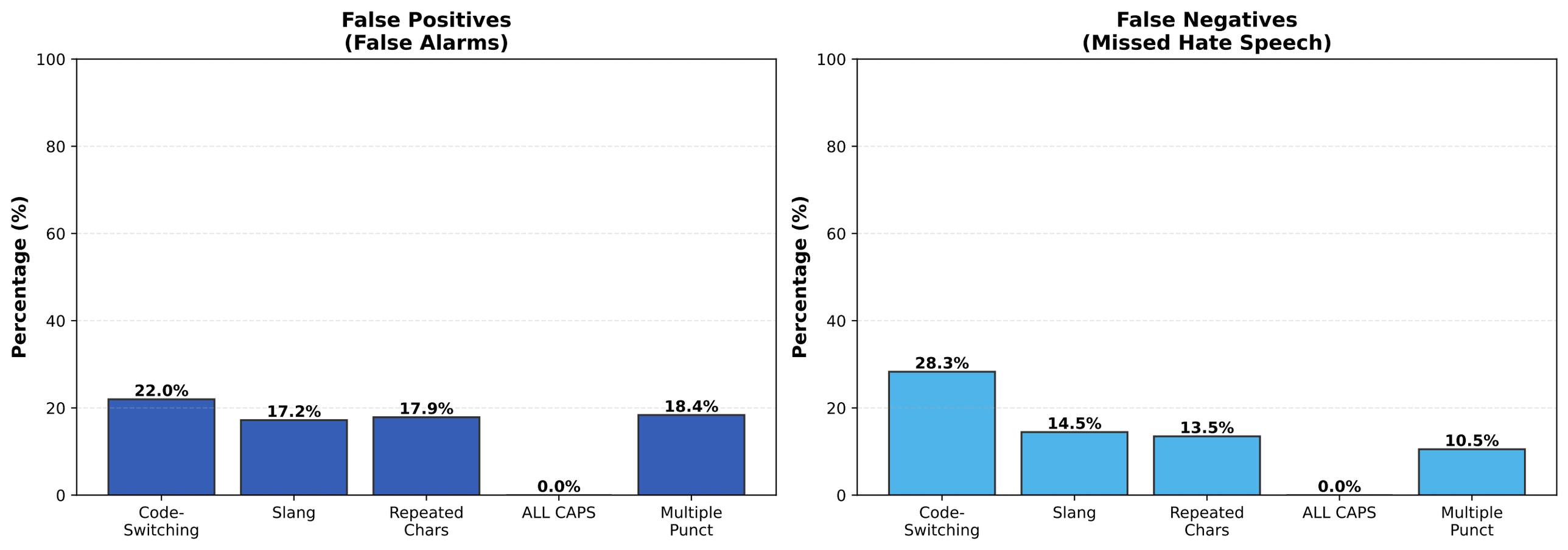
False Positives
(False Alarms)
Percentage (%)
100
80
60
40
20
0
22.0%
17.2%
17.9%
0.0%
18.4%
Code-
Switching
Slang
Repeated
Chars
ALL CAPS
Multiple
Punct
False Negatives
(Missed Hate Speech)
Percentage (%)
100
80
60
40
20
0
28.3%
14.5%
13.5%
0.0%
10.5%
Code-
Switching
Slang
Repeated
Chars
ALL CAPS
Multiple
Punct