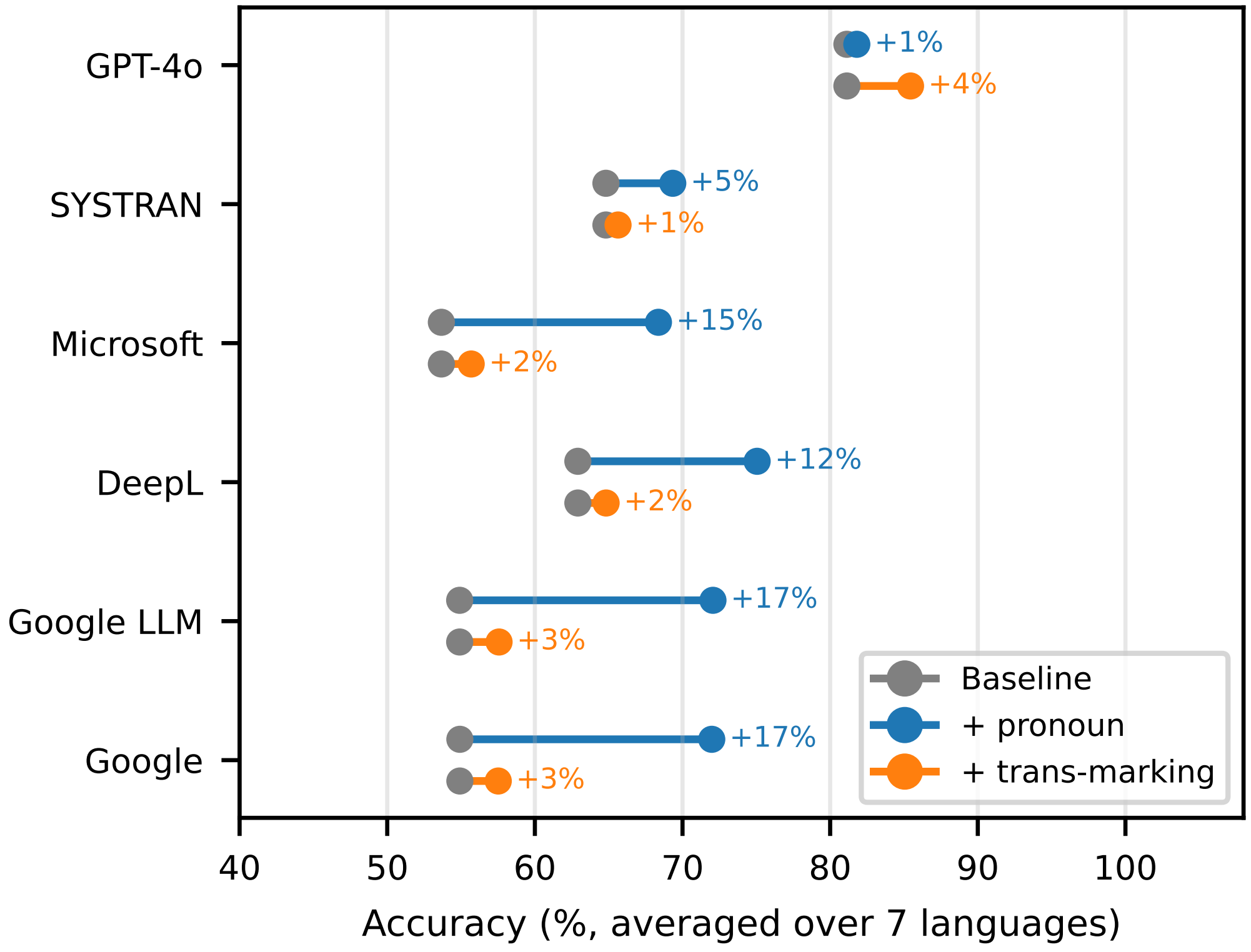

GPT-4o
SYSTRAN
Microsoft
DeepL
Google LLM
Google
+1%
+4%
+5%
+1%
+15%
+2%
+12%
+2%
+17%
+3%
+17%
+3%
Baseline
+ pronoun
+ trans-marking
40
50
60
70
80
90
100
Accuracy (%, averaged over 7 languages)