Controlled Comparison: Fixed-k Baselines vs RL Dynamic TopK

Legend:
- No retrieval
- Fixed-k baselines
- Best fixed-k
- RL Dynamic TopK

F1 Score (%)

- k=0 (no retrieval): 18.5%
- k=1: 31.8%
- k=3: 34.1%
- k=5: 34.4%
- k=1: **34.9%**
- RL Dynamic TopK: **35.1%**

avg k=7.0

30% fewer retrievals vs k=10