GPT-2 CODI: Token Importance vs Attention Distribution (Layer 11)
Total attention to CoT tokens: 11.66% | Tokens 2 & 3 critical but attention uniform

Legend:
- Importance (% accuracy drop when ablated)
- Attention (% of attention among 6 CoT tokens)

Token Position	Importance	Attention
Token 0	21.4%	17.6%
Token 1	21.4%	17.2%
Token 2	50.0%	16.8%
Token 3	50.0%	16.5%
Token 4	19.0%	16.1%
Token 5	23.8%	15.8%