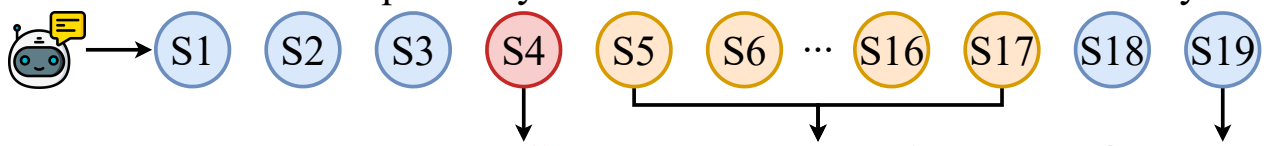
Task: What writer is quoted by Merriam-Webster for Word of the Day from June 27, 2022?



S4 Root Cause: web_request ".../jingoism-2022-06-26...". Date should be 2022-06-27.

① Parse: Unified Trace Parsing

Step sequence: each step (agent name, input/output, tools called)
Agent dependency: agent and agent, agent and tool
System configuration: prompt and tool schemas

agent: ck_plan_agent, ck_action_agent, web_plan_agent...
tool: simple_web_search, web_agent, file_agent...

② Evaluate: Evidence-based Evaluation

✦ **Global Level**
Metrics❌: Tool Correctness, State Consistency
Step-level Evidence: S4, Tool Param Value Error, date should be 06-27
✦ **Agent Level (ck_action_agent, S4-S17)**
Metrics❌: Tool Correctness, Reasoning Consistency
Step-level Evidence: S4, Tool Param Value Error, date error in request

③ Extract: Principle Extraction

✦ **Execution Principle (Agent Behavior)**
- Title: "Validate URL Date Parameters Before Navigation"
- Content: When calling web_agent with date-specific URLs, verify date matches task requirement...
- Source metrics: [Tool Correctness, State Consistency]

✦ **Judge Principle (Evaluate Standard)**
- Title: "Detect URL Parameter Mismatches"
- Content: When evaluating Tool Correctness, check if URL params match task. Mark as Tool Param Value Error if violated.
- Source metrics: [Tool Correctness]

④ Manage: Principle Management

✦ **Principle Bank Management**
- GroupBy: type (execute/judge), level (global/agent), agent
- Complete-Linkage Clustering (adaptive threshold)
- LLM Refine: add, modify, merge
- Principle Format: {"principle_id": "f8a2b1c3d4e5...",
"function": "execute", "type": "global", "agent_name": null,
"title": "Validate URL Date Parameters Before Navigation",
"content": "When calling web_agent with date-specific...",
"embedding": [0.123, -0.456...], "source_cases": ["518836..."] }

✦ **Principle Retrieval and Prompt Augmentation**
- Principle Retrieval: Embedding + Keyword Search → LLM Rerank
- Prompt Augmentation:
<principles>Principles are advisory, not mandatory.
global level
title: Validate URL Date Parameters Before Navigation
content: When calling web_agent with date-specific URLs, verify date matches task requirement...</principles>

✦ **Bidirectional Self-Evolution**
Execution Principle→Agent behavior ↑
Judge Principle→Evaluate Standard ↑

⑤ Verify: Fault Injection

- Input: Give Successful trace and specify target fault type.
✦ **LLM selects injection step and modifies step to satisfy binary check criterion and propagates error through subsequent steps.**
- Output: Ground truth label
{"fault_type": "Tool Param Value Error", "root_cause_step": 35,
"original_correct_answer": "All tests passed!","wrong_final_answer": "Unable to verify..."
"causality_chain": "step35: str_replace_editor called with command='invalid_command' → tool returns constraint violation error → agent cannot run verification test → wrong final answer"}
Enables: type accuracy (correct fault type) and localization accuracy (correct root cause step).