

Productive Knowledge of Chinese Character Form Patterns in Human Memory and LLMs

This study investigates the psychological reality and productivity of Chinese character grammar by testing whether internalized structural regularities facilitate memory for novel logographic forms and comparing these patterns to Large Language Models (LLMs). While Chinese orthography is often characterized as a system requiring rote memorization of thousands of discrete logograms, research in grapholinguistics suggests that readers actually internalize a productive character grammar. This grammar allows for the systematic evaluation of brand-new visual forms based on learned constraints regarding stroke placement, radical position, and configuration (Myers, 2019). To test this, we used stimuli from the Khitan scripts. In the first task, 32 native Chinese speakers provided likeness judgments for 100 Khitan characters. The results yielded a robust replication of the structural likeness effects found in Myers (2025) ($r = .803, p < .001$), confirming that readers possess consistent, graded intuitions about which forms are well-formed according to Chinese structural principles.

In the second task, we conducted a recognition memory experiment to determine if this productive structural knowledge facilitates encoding. Participants were tested on their ability to recognize the Khitan forms after a short delay. While initial analysis showed a non-significant trend between structural likeness and memory accuracy ($p = .181$), a closer examination of response patterns suggests that structural familiarity primarily influences response bias rather than overall mnemonic sensitivity (d'). Specifically, characters with high likeness scores were more likely to be judged as "old," regardless of whether they had appeared in the training set. These results imply that the 5-minute retention interval may have been insufficient to capture the full consolidation effects of orthographic grammar on memory.

Finally, when LLMs were tasked with evaluating the same Khitan stimuli, the models exhibited a significant ceiling effect that assigned high acceptability scores to nearly all items. This demonstrates a critical divergence between human and LLMs. While LLMs effectively capture broad statistical regularities of Chinese forms, they appear to lack the inhibitory grammar that human readers use to strictly reject malformed structures. These findings suggest that human orthographic productivity is defined as much by its constraints and rejections as by its permissive patterns. The permissive nature of LLMs reflects a statistical sensitivity that does not yet mirror the "picky" structural integrity of human linguistic intuition. This research underscores the need for more rigorous comparative methods to bridge the gap between human psycholinguistic reality and artificial pattern recognition.

Keywords: Character grammar; Khitan script; Productive orthographic knowledge; Recognition memory; Large language models