

# Scaling Limits of the UM KN Chain: A Synthesis of 2026-03-13 Findings

Claude and MJC

2026-03-13

## Abstract

We report four interlocking negative results about the UM’s KN chain scoring system. Together, they show that the UM’s cascading threshold architecture cannot match external KN-6 at any scale, and that the gap *grows* with data size. The UM’s value lies not in raw compression but in its interpretable memory traces. We propose a split architecture: external KN for compression, UM for structure and interpretation.

## 1 Four Negative Results

### 1.1 Scoring Protocol Error

Previous work reported that the UM’s frozen KN score (2.628 bpc at 100K) beat external KN-6’s online score (2.719). This comparison was invalid: it compared frozen (score with future data) against online (predict-then-observe).

Under matched protocols:

Protocol	UM	KN-6	Gap
Online vs online (100K)	3.203	2.719	+0.484
Frozen vs frozen (100K)	2.628	1.265	+1.363

The UM is worse under both protocols. Frozen scoring inflates the UM’s advantage because high-order contexts with few observations benefit disproportionately from full-data counts.

### 1.2 Coverage Bottleneck

The UM creates higher-order neurons through cascading threshold: a  $k$ -gram neuron requires its  $(k-1)$ -gram parent to exist and fire. At 100K with  $\tau=1$ :

ES	UM neurons	KN-6 contexts	Coverage
bigram_prev	2252	2252	100%
trigram_prev	2252	2252	100%
4gram_prev	2111	21336	10%
5gram_prev	1607	35003	4.6%

At orders 3–4, coverage is complete because the parent ESs are pre-populated (byte\_prev, bigram\_prev). At orders 5–6, coverage drops precipitously because parent neurons must be dynamically created and only a fraction of contexts cascade through all levels.

### 1.3 Threshold Scaling Degradation

Lowering  $\tau$  to 1 creates neurons for every observed context, matching coverage at orders 3–4. But at orders 5–6, sparse contexts *degrade* performance as scale increases:

Scale	UM $\tau=1$	UM $\tau=4$	KN-6 frozen	Gap $\tau=1$
100K	2.605	2.749	1.265	+1.340
300K	2.778	2.813	1.272	+1.506
500K	2.920	2.890	1.291	+1.629
1M	3.101	3.054	1.335	+1.766

Both  $\tau$  settings degrade:  $\tau=1$  from 2.605 to 3.101,  $\tau=4$  from 2.749 to 3.054. At 500K,  $\tau=4$  overtakes  $\tau=1$  (the crossover persists at 1M). KN-6 is roughly flat (1.265  $\rightarrow$  1.335). Root cause:  $\tau=1$  creates many neurons at high orders, each with 1–2 observations. KN discount from these sparse contexts steals probability mass.  $\tau=4$  degrades more slowly because it requires 4 observations before creating a context.

### 1.4 Multi-Pass Diminishing Returns

Multiple retroactive passes (UMR\_RETRO\_PASSES) and multiple online passes (UMR\_ONLINE\_PASSES) provide marginal improvement:

Passes	100K frozen	200K frozen	Gain
1 retro	2.628	2.610	(baseline)
3 retro	2.605	2.599	-0.023
5 retro	2.600	—	-0.028
10 retro	—	2.592	-0.018

Each additional pass captures roughly 40% of the previous gain. The convergence is fast because the bottleneck is structural (coverage, count ratios) not observational (missing data).

## 2 Root Cause Analysis

The four results share a common root: the UM’s cascading threshold creates *structurally different* count distributions than external KN.

External KN stores all  $n$ -gram contexts in a flat hash table. Each context’s count is proportional to its frequency in the data. The KN discount  $D \approx 0.9$  is calibrated to these frequencies.

The UM stores counts in LPP entries attached to neurons. Neurons are created through cascading threshold, which introduces two distortions:

1. **Missing contexts:** Only contexts whose parent neuron exists and fires can be observed. At order 6, 95.4% of KN contexts are missing from the UM.
2. **Front-loaded creation:**  $\tau=1$  creates entries at first observation. The entry exists from position  $p$  onwards, but external KN has counts from position 0. Retroactive passes partially fix this but can’t create entries for contexts whose parents don’t exist.

## 3 Implications and Path Forward

The UM KN chain is a structurally limited implementation of external KN. No parameter tuning ( $\tau$ , order, passes) can close the gap because the limitation is architectural.

However, the UM provides something external KN cannot: **interpretable memory traces**. At every position, the UM produces a boolean algebra sentence explaining its prediction—which source events contributed, through which LPPs, with what weights. This is the UM’s native reasoning made visible.

We propose a split architecture:

- **Compression:** External KN-6 (already working at 1.925 bpc for 100M via `umr compress mode 0`). Add sparse bigrams and match copying for further gains.
- **Interpretation:** UM memory traces explain predictions, identify structural patterns (word boundaries, tag structure, numeric sequences), and provide provenance for errors.
- **Factor discovery:** Use UM’s event space architecture for Track 2 (lexicon, phrase, sentence structure). The UM’s advantage is in discovering hierarchical structure, not in raw byte prediction.

**Evidence for factor value.** The word-kn6 experiment shows that word-level structure improves KN-6 at multiple scales:

Scale	KN-6	+Word mix	Gain	Top-100 gain
100K	2.719	2.608	+0.111	+0.242
1M	2.397	2.338	+0.060	+0.113

The gain decreases with scale ( $0.111 \rightarrow 0.060$ ) as KN-6’s high-order contexts capture some word structure implicitly. But it remains positive and concentrated at word boundaries. The multi-frequency model adds tag context for +0.184 bpc at 1M (KN-6  $2.398 \rightarrow 2.214$ ). These gains come from *higher-level structure*—exactly what the UM’s factor discovery architecture is designed to capture. The gap from KN-6+sparse+match (1.588 bpc at 1B) to the Hutter Prize record ( 0.93 bpc) is 0.66 bpc, which plausibly decomposes into word/phrase/sentence/semantic structure.