Scoring Protocol Correction:
UM vs External KN-6

Claude and MJC

March 2026

Abstract

We correct an error in the order scaling analysis. The claim that "the UM beats external KN-6" compared UM *frozen* scoring (predict after seeing all data) against KN-6 *online* scoring (predict before observing). These are different protocols. Under matched protocols, the UM is worse at all scales.

1 The Error

The previous order scaling paper reported:

	UM frozen	KN-6 online	Δ
100K, $\tau = 1$	2.628	2.719	-0.091 ("UM wins")

This comparison is invalid. Frozen scoring uses counts from the *entire dataset* at every position, while online scoring uses only counts from positions before the current one. Frozen scoring always gives lower (better) bpc on the training data.

2 Correct Comparison

External KN-6 frozen scoring: count all data once, then score.

Order	Online bpc	Frozen bpc
1 (unigram)	3.848	3.723
2 (bigram)	3.124	2.708
3 (trigram)	2.756	1.942
4	2.674	1.546
5	2.691	1.364
6	2.719	1.265

Frozen scoring is dramatically better: -1.454 bpc at order 6 (100K). Higher orders benefit more from frozen scoring because there are more contexts to exploit.

3 UM Under Both Protocols

	UM	KN-6	Gap
Online (pass 1)	3.203	2.719	+0.484
Frozen (pass 3)	2.628	1.265	+1.363

The UM is worse under both protocols. The frozen gap (+1.363) is larger because:

1. **Coverage**: The UM creates neurons for only 23% of KN-6's contexts (order 5: 10%, order 6: 4.6%).

2. **Scoring**: The UM's KN chain uses the same interpolation as external KN, but with fewer contexts.

4 Why Frozen Helps the UM Less

KN-6 frozen at 100K: 1.265 bpc (-1.454 from online). UM frozen at 100K: 2.628 bpc (-0.575 from online 3.203).

The UM gains less from frozen scoring because it has fewer high-order contexts. Frozen scoring helps most at high orders (where rare contexts provide specific predictions). With only 4.6% coverage at order 6, the UM can't exploit frozen counts for the missing 95.4% of contexts.

5 Scaling at 1M

	Online bpc	Frozen bpc
KN-1	3.882	3.840
KN-2	3.091	2.926
KN-3	2.583	2.185
KN-4	2.383	1.717
KN-5	2.366	1.477
KN-6	2.397	1.335

Frozen gains at 1M (-1.062 at order 6) are smaller than at 100K (-1.454) because online counts are already large at 1M.

6 Implications

For compression. The UM's frozen score cannot be used for compression because it requires seeing the data before scoring. For lossless compression, only the online protocol is valid (or a block protocol where each block is trained on previous blocks).

For the UM. The retroactive advantage is an artifact of comparing different protocols. Under matched protocols, the UM is strictly worse than external KN-6 at all scales tested.

The path forward. Close the gap by: (1) increasing context coverage to 100% at all orders, and (2) using online scoring for compression.