Correlation Dimension of Auto-Regressive Large Language Models

Xin Du and Kumiko Tanaka-Ishii

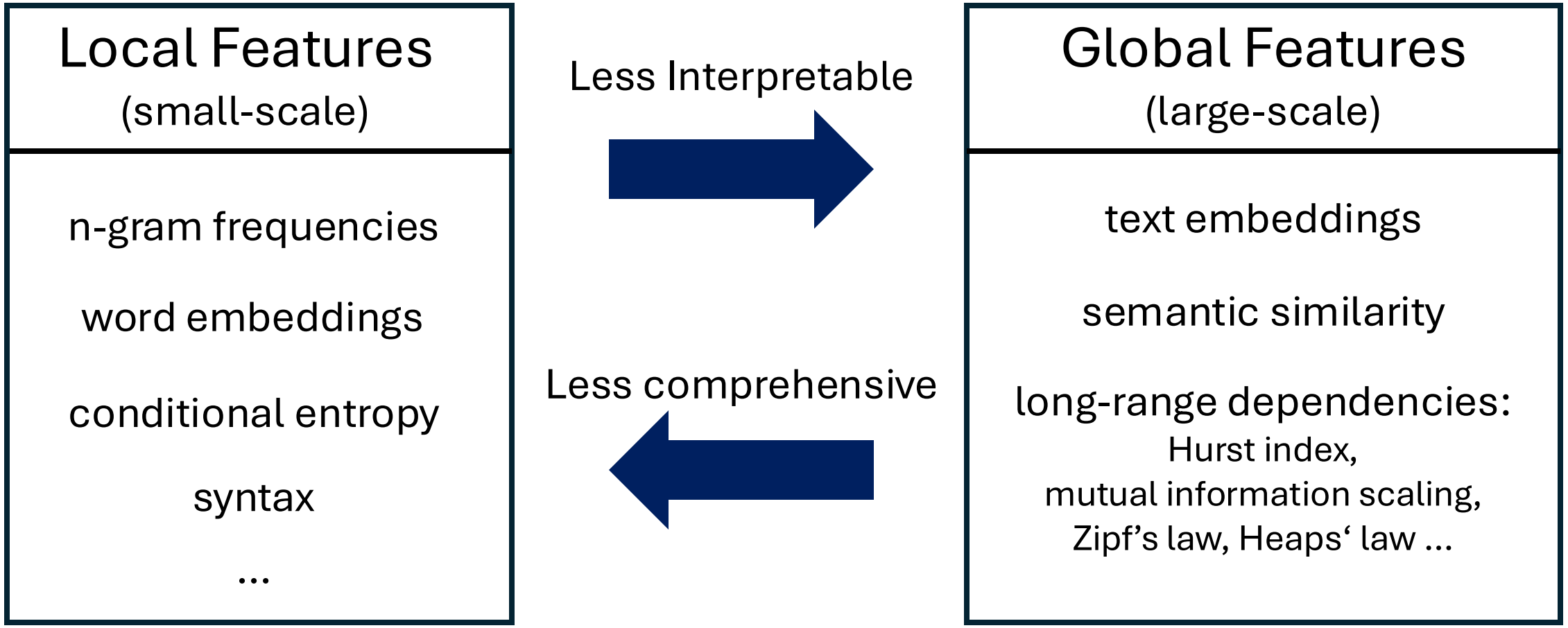
Waseda University

arXiv

Homepage

NEURAL INFORMATION PROCESSING SYSTEMS 2025

Gap Between Local & Global Text Features

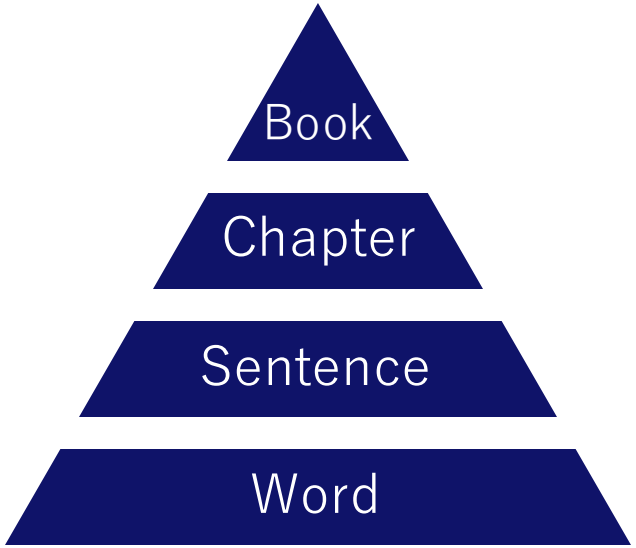


Need a new metric

that is locally interpretable and globally comprehensive

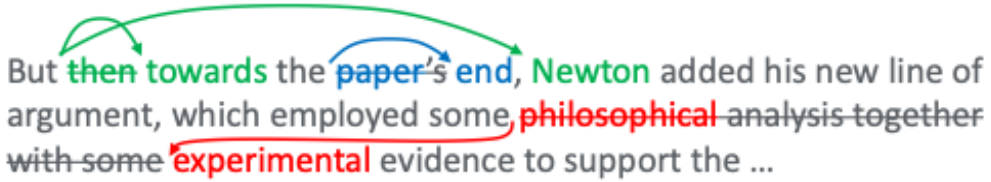
CorrDim: Bridging the Gap via Self-Similarity Across Scales

Self-similarity (complexity science) naturally occurs in natural language
(and many other real-world systems)



Correlation dimension monitors recurrences in text generation process

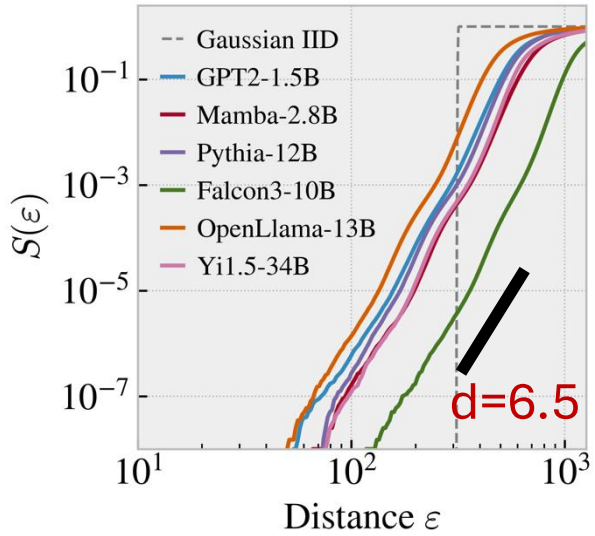
Recurrence of an LLM state: skippable text segment



$$S(\varepsilon) \propto \varepsilon^{d}$$

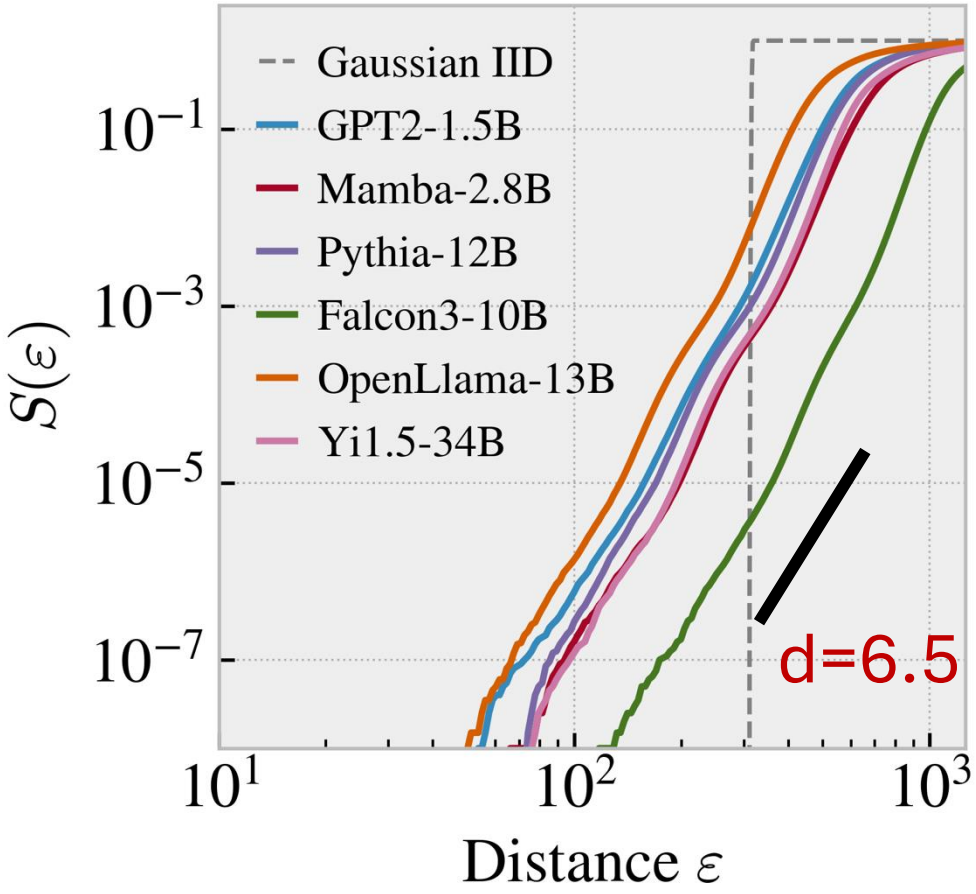
# of recurrences

Distance threshold

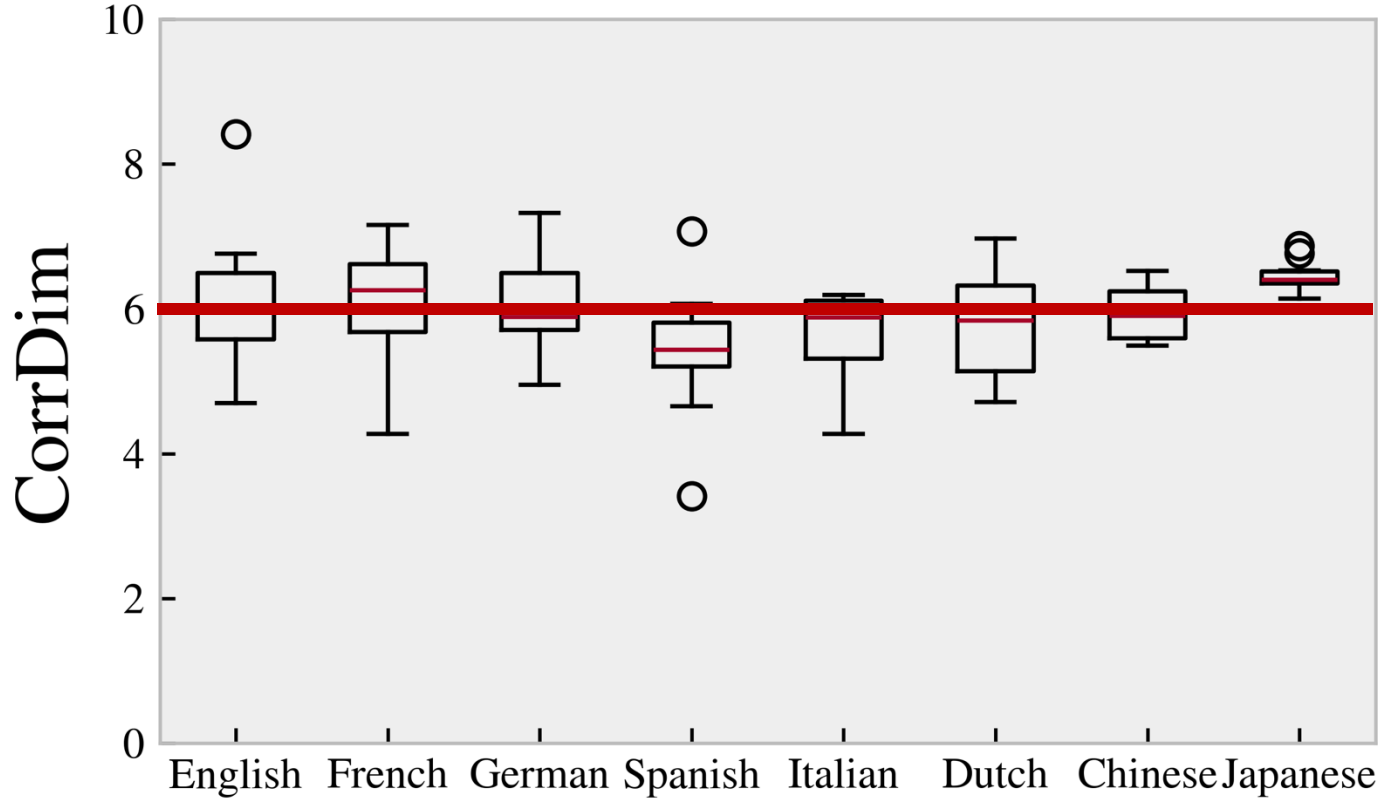


CorrDim ≈6 for natural languages
≈5 for programming languages

Consistency across
LLM models, architectures:



Consistency across languages

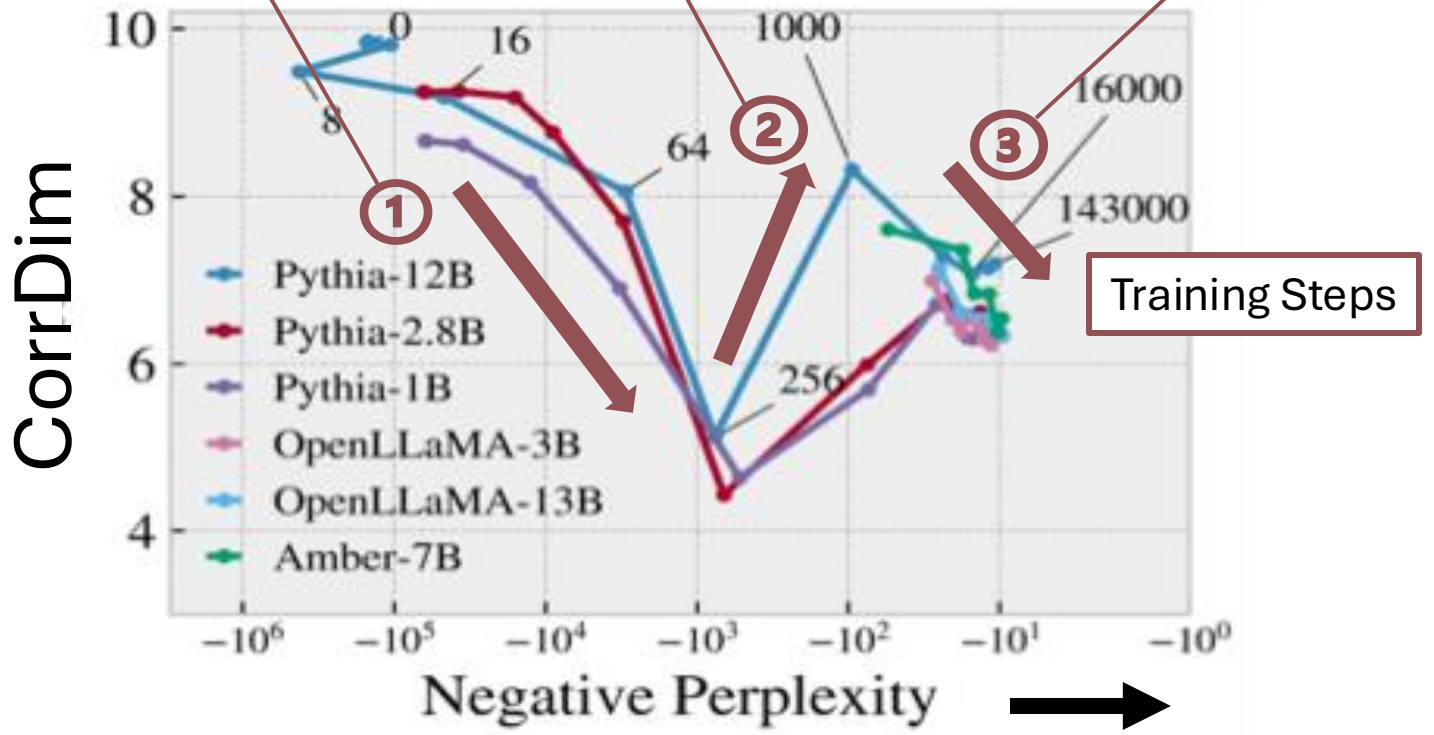


CorrDim transition reveals **3 stages** in LLM pretraining

① Learning of **short-range** dependency (n-gram statistics)

② Learning of **long-range** dependencies

③ **Generalization** or “forgetting”

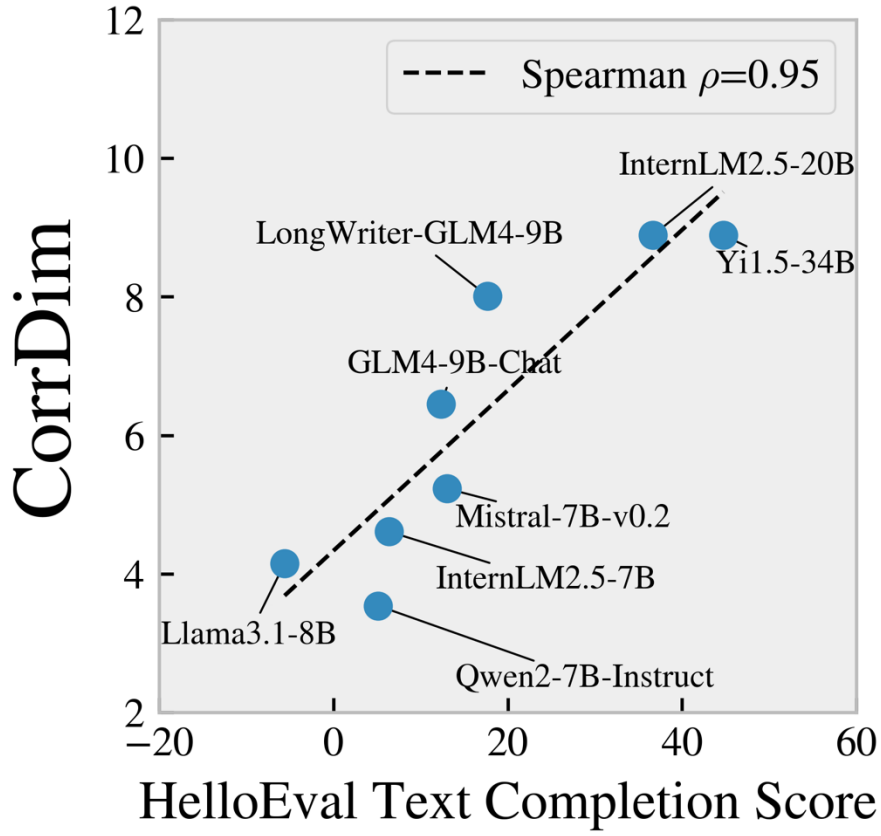


CorrDim indicates curious macroscopic LLM generation behaviors

Degeneration:

generation becomes repetitive, incoherent, bland

High correlation (ρ=0.95)
with hand-crafted LLM benchmark



Hallucination:

generating unfactual texts

LLM showing low CorrDim on
knowledge-intensive texts risks hallucination

Model	Normal text (ave.)	Knowledge-intensive text	Recalling or Hallucinating
Qwen2.5-0.5B	5.88	3.32	hallucinate
Qwen2.5-7B	6.27	3.56	hallucinate
Qwen2.5-32B	6.32	4.42	hallucinate
Falcon3-1B	6.03	3.28	hallucinate
Falcon3-3B	6.11	3.14	hallucinate
Falcon3-7B	6.55	6.68	recall
Falcon3-10B	6.56	8.49	recall

drop

rise

Qwen2.5-7B (CorrDim: 3.56):

Robert Schofield, Robert McAfee Brown, Kei… ohn H. Cobb, Jr., Arthur Roberts , Robert Solomon… Bruce L. Reichen-bach .

fabricated names

Falcon3-10B (CorrDim: 8.49):

Delwin Brown, David Ray Griffin, Hank Keeton, Georg… R. Reynolds, Franklin I. Gamwell, Alan I. Megill, Roland Faber, an… Keller.

correct names

Reason: Hallucination requires far less complexity in generation.

Summary

We propose CorrDim, a new metric from self-similarity and dynamical systems theory.

- CorrDim captures both local & global features, based on self-similarity
- CorrDim quantifies macroscopic behaviors of LLMs, including degeneration and hallucination
- CorrDim is efficient, requiring zero memory overhead, computed inplace during inference.

Xin Du.
duxin@aoni.waseda.jp

Kumiko Tanaka-Ishii.
kumiko@waseda.jp