Computational Humanities Research

www.cambridge.org/chr

Cognitive stylometry: A computational study of defamiliarization in modern Chinese

Maciej Kurzynski

Department of Chinese, Lingnan University, Hong Kong

Research Article

Cite this article: Kurzynski Maciej. 2026. "Cognitive stylometry: A computational study of defamiliarization in modern Chinese" *Computational Humanities Research*, **2**:e1, https://doi.org/10.1017/chr.2025.10020

Received: 31 December 2024
Revised: 5 August 2025
Accepted: 4 November 2025

Keywords:
Chinese literature; information theory; large language models; perplexity; predictive coding

Corresponding author:
Maciej Kurzynski;
Email: maciej.kurzynski@ln.edu.hk

Abstract

Autoregressive language models generate text by predicting the next word from the preceding context. The regularities internalized from specific training data make this mechanism a useful proxy for historically situated readerly expectations, reflecting what earlier linguistic communities would find probable or meaningful. In this article, I pre-train a GPT model (223M parameters) on a broad corpus of Chinese texts (*FineWeb Edu Chinese V2.1*) and fine-tune it on the collected writings of Mao Zedong (1893–1976) to simulate the evolving linguistic landscape of post-1949 China. Identifying token sequences with the sharpest drops in perplexity – a measure of the model's surprise – reveals the core phraseology of "Maospeak," the militant language style that developed from Mao's writings and pronouncements. A comparative analysis of modern Chinese fiction demonstrates how literature becomes unfamiliar to the fine-tuned model, generating perplexity spikes of increasing magnitude. The findings suggest a mechanism of attentional control: whereas propaganda backgrounds meaning through repetition (cognitive overfitting), literature foregrounds it through deviation (non-anomalous surprise). By visualizing token sequences as perplexity landscapes with peaks and valleys, the article reconceives style as a probabilistic phenomenon and showcases the potential of "cognitive stylometry" for literary theory and close reading.

Plain language summary

This study applies computational methods to revisit the Formalist theory of defamiliarization, which posits that art renews perception by rearranging familiar forms in unpredictable ways. The research goal was to formalize and measure the stylistic tension between predictability and surprise using a large language model as a stand-in for a historically situated reader. To achieve this, the study employed a multi-stage experimental design. First, a GPT-2 style transformer model (223M parameters) was pre-trained on a large, high-quality corpus of modern Chinese text (FineWeb Edu Chinese V2.1) to establish a neutral linguistic baseline. This model was then fine-tuned for five epochs exclusively on the *Selected Works of Mao Zedong* to simulate a reader becoming saturated with "Maospeak," the highly formulaic and militant language of the Mao era. The core metric was perplexity, which quantifies the model's predictive uncertainty, or "surprise," at each token. The primary finding was a dual process of familiarization and defamiliarization. By tracking the largest decreases in perplexity, the study identified the core phraseology of Maospeak, including canonical lists of class enemies and rigid rhetorical patterns. This "cognitive overfitting" demonstrated how propaganda automates language to create a predictable, low-perplexity environment and legitimatize violence. *N*-gram entropy analysis confirmed that the Mao corpus was measurably more repetitive than a literary corpus at smaller *n*-gram sizes, i.e., in its foundational vocabulary and short collocations. Concurrently, as the model specialized in Maospeak, its average perplexity on a corpus of 100 modern Chinese novels increased, making literary language appear more alien. Microscopic analysis of novels by Zhang Wei, Lilian Lee and Dung Kai-cheung revealed how literary style creates "non-anomalous surprise" through high-perplexity moments, such as creative metaphors, the use of classical Chinese syntax and the abrupt insertion of non-standard Chinese languages such as Cantonese. Ultimately, the research proposes "cognitive stylometry" as a new mode of inquiry, redefining style as a measurable cognitive signature that orchestrates a reader's predictive processing by balancing familiar grounding with moments of meaningful deviation.

Introduction: Rethinking defamiliarization in the age of LLMs

Large language models (LLMs) present an exciting opportunity for literary theory, inviting new perspectives on the foundational concepts, such as meaning, attention and memory. Yet, in addition to ethical concerns (Weidinger et al. 2021; Hendrycks, Mazeika, and Woodside 2023; Kirschenbaum 2023; Yan et al. 2024), their integration within humanistic studies faces two disciplinary challenges. On the one hand, radically anthropocentric beliefs continue to frame machine intelligence through a lens of inadequacy, as seen in Noam Chomsky's critique

This article was awarded Open Data and Open Materials badges for transparent practices. See the Data availability statement for details.

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (https://creativecommons.org/licenses/by-nc/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

of ChatGPT's "false promise" (Chomsky 2023) or Michele Elam's assertion that the "creative process cannot be reduced to a prompt" (Elam 2023). On the other hand, computational approaches that *are* adopted by literary scholars risk veering so far into benchmarking and quantification that they lose sight of literature as a source of aesthetic experience and interpretive engagement. These concerns are particularly relevant to the computational analysis of literary style, or stylometry, where the focus on precision metrics often overshadows the very texts under study. Paradoxically, highly accurate attribution models might fail to tell us anything about style.

An alternative path may lie in restarting the conversation between probabilists and humanists that unfolded in the early twentieth century but then subsided under historicist pressures. In 1913, Andrey Markov analyzed 20,000 letters from Pushkin's poem *Eugene Onegin* to demonstrate his theory of dependent trials, later known as Markov chains, proving that the probability of a letter being a vowel or a consonant depends on the class of the preceding letter (Markov 2006). Decades later, Claude Shannon reframed this problem in the empirical terms of information theory: "How well can the next letter of a text be predicted when the preceding *N* letters are known?" (Shannon 1951, 50). He demonstrated that humans possess an implicit knowledge of language statistics, effectively casting them as predictive language models. The aesthetic implications of this tension between predictability and surprise were anticipated by the Formalist scholar Viktor Shklovsky, who argued that art combats the "automatization of perception," a process whereby, through habit, life "becomes nothing and disappears" (Shklovsky 2017, 80). By contrast, in a process he called defamiliarization (*ostranenie*), art "increases the complexity and duration of perception" and "recovers the sensation of life" by rearranging familiar forms in unpredictable ways (Resseguie 1991; Steiner 1984).

Recent advances in computational hardware, artificial intelligence and cognitive science call for a renewed humanistic perspective on these complementary phenomena. Trained on vast corpora spanning literature, journalism and social media, LLMs represent what Shklovsky could call the "familiar," a statistically distilled reflection of human knowledge, language use and cultural expectations. Meanwhile, the concept of "defamiliarization" has been reinvigorated with new empirical evidence, from eye-tracking experiments showing that readers fixate longer on unexpected words (McDonald and Shillock 2004; Bonhage et al. 2015) to fMRI scans revealing heightened mental engagement with unusual sentence structures, consistent with theories of the brain as a predictive organ (Dehaene et al. 2005; Schuster et al. 2025; Henderson et al. 2016; Bohrn et al. 2012; Clark 2013, 2016). Shannon's blurring of the human/machine reader distinction has also been pushed further, with research suggesting parallels between narrative comprehension in humans and in Transformer-based models (Caucheteux and King 2022; Kumar et al. 2024), though scholars urge caution (Feghhi et al. 2024; Soni et al. 2024; Fodor, Murawski, and Suzuki 2025). Humanists are particularly well-positioned to contribute to this interdisciplinary dialogue – not to surrender to the discourse of technoscience, but to offer new perspectives on literature as a source of multi-layered engagements with our cognitive, emotional and predictive faculties (Kurzynski 2025), a goal that motivates this project.

Methods: Predictability and perplexity

The tension between expectation and surprise, central to both information theory and the notion of defamiliarization, can be formalized using perplexity, a measure of a statistical model's predictive certainty (Jelinek et al. 1977). Lower perplexity on the correct output indicates high confidence, while higher perplexity signals surprise. Most of the currently available LLMs learn to minimize this surprise by iterating over large amounts of text, either by predicting the next token in a sequence (autoregressive models) or by reconstructing an original sequence from a perturbed one (masked language models) (Paaß and Giesselbach 2023, 19–78).

In this study, perplexity will serve as a metric for simulating human engagement with texts. While acknowledging the significant differences between natural cognition and artificial intelligence, notably the different ways in which they emerge (Lillicrap et al. 2020), I leverage language models as proxies of historically situated human readers to revisit Shklovsky's argument (Varnum et al. 2024). Style, in this framework, is not merely a property of the text itself but is understood in relation to the predictive coding mechanisms inherent to both biological and artificial cognition. Phrases that elicit high perplexity will be especially relevant here, aligning with the understanding of style as a "deviation from the norm" (Enkvist, Spencer, and Gregory 1964; Herrmann, Dalen-Oskam, and Schöch 2015), whereas highly predictable phrases will instead reveal the automatized linguistic background against which stylistic distinctions emerge. Of particular importance to this article, however, is how an initially unfamiliar phrase becomes familiar through repeated exposure, and, conversely, how a particular phraseology becomes surprising when a model overfits on a different linguistic style. Such a process may offer a compelling computational analog for phenomena of broader cognitive-historical interest: for instance, how literary canons form, or how, under specific historical circumstances, prescriptive linguistic norms violently narrow the scope of legitimate expression (Bloch 1975; Epstein 1990; Ji 2004; Hughes 2009; Klemperer 2013).

Stylometrists have rarely considered predictability a primary stylistic feature, traditionally favoring more tangible elements like function words, syntactic structures and thematic profiles (Lutosławski 1898; Mosteller and Wallace 1963; Burrows 1987; Herrmann, Dalen-Oskam, and Schöch 2015; Rybicki, Eder, and Hoover 2016; Evert et al. 2017). This has changed with the advent of LLMs, whose human-like output spurred an urgent search for markers to differentiate machine from human text, with perplexity emerging as a key candidate. An influential early paper by Holtzman et al. argued that human text is not the most probable text: "natural language rarely remains in a high probability zone for multiple consecutive time steps, instead veering into lower-probability but more informative tokens" (Holtzman et al. 2020). Subsequent studies confirmed that low-perplexity sequences often indicate machine provenance (Mitchell et al. 2023; K. Wu et al. 2023). However, as newer LLMs produce text that is "nearly indistinguishable from human-written text from a statistical perspective" (Chakraborty et al. 2023), researchers have developed methods like watermarking, which embeds statistically detectable patterns into AI-generated content for provenance verification (Zhao et al. 2023; Kirchenbauer et al. 2023). A few studies in the computational humanities have also begun to explore perplexity's stylometric potential, using it to distinguish canonical from non-canonical novels (Y. Wu et al. 2024), machine from human translation (Bizzoni et al. 2020) and political discourse from novelistic prose (Kurzynski 2023, 2024), signaling a broader turn toward evidence of cognitive and informational processing in literary analysis.

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

Datasets and tokenizers

The foundation of this study is the `Fineweb Edu Chinese V2.1`, a large-scale, high-quality Chinese corpus constructed from diverse datasets, including *Wudao*, *Skypile*, *IndustryCorpus* and *Tele-AI* (Yu et al. 2025). The *Fineweb* authors employed a filtering pipeline where a scoring BERT model evaluated each document on a 0-to-5 scale for educational value and writing quality. The present study utilizes the "4–5" split, which contains only the highest-scoring, deduplicated content. From this top-tier partition of approximately 70 GB, one million Chinese passages varying in length from 256 to 1,024 characters were sampled for pre-training, serving as a neutral linguistic environment. Passages in languages other than Chinese as well as those with a high percentage of numbers and scientific notation were removed.

Two distinct test corpora were then constructed to represent stylistically opposed domains. The first is a seven-volume compilation of the *Selected Works of Mao Zedong* 毛泽东选集, which combines the official five-volume collection (covering 1925–1957) with the two unauthorized "Jinghuo" 静火 volumes, extending the collection through 1976. Together, these texts comprise Mao's most significant speeches, essays and directives, establishing the core tenets of Mao Zedong Thought and representing the target style for the familiarization hypothesis (Leese 2014). The second test corpus is the *Top 100 Chinese Novels of the 20th Century*, a canon compiled in 1999 by *Asia Weekly* 亚洲周刊 in tandem with a jury of fourteen literary experts (Qiu 1999). The Chinese term *xiaoshuo* 小说 (lit. "small talk") is broader than the English "novel," so this list includes both full-length novels and short story collections from Mainland China, Taiwan and Hong Kong, providing a diverse sample of modern literary styles. All texts were normalized and converted to simplified Chinese. To prevent longer documents from dominating the datasets, a sampling technique was applied to ensure that each document from either corpus could contribute no more than 100 chunks of 1,024 characters (Table 1).

To prevent data leakage, i.e., the inclusion of evaluation content in the training data, an *n*-gram hashing technique was employed. This is crucial for measuring familiarity with a style rather than memorization of specific passages or even entire documents. All documents in the test corpora were processed to generate a set of unique, character-level 13-grams. This *n*-gram size is long enough to capture verbatim quotations but short enough to avoid falsely flagging common Chinese expressions; it also aligns with contamination analysis methods used for LLMs (Brown et al. 2020; OpenAI 2023). Each 13-gram was hashed using MD5 to create compact fingerprints, which were then used to scan all training documents. Any match was verified with a literal string search, and confirmed documents were discarded from the training set.

The conversion of raw text into numerical identifiers via a tokenizer is a foundational step that shapes the outcome of any computational analysis of language. While state-of-the-art models frequently employ subword tokenizers like byte-pair encoding (BPE), these algorithms introduce significant analytical challenges, particularly for logographic languages like Chinese that lack explicit word delimiters. For example, BPE tokenizers operate by iteratively merging frequent pairs of characters or subwords, which can result in segmentations that are both non-intuitive and inconsistent. A greedy BPE tokenizer might segment 学科技 ("to study electronics") as 学科 ("discipline") 技 ("skill"), rather than 学 ("to study") 科技 ("electronics"). This happens because 学科 ("discipline"), an otherwise valid word, has been learned first by the tokenizer during training and thus appears earlier in the list of merges. Similarly, 第二天 ("the next day") might be merged into a single token, while 第三天 ("the third day") might not, as the former occurs more frequently in the training data, etc.

While such inconsistencies may not drastically alter aggregate perplexity scores across a large corpus, they can severely distort the close analysis of individual sequences, mistaking the quirks of a tokenizer for genuine authorial choices. A BPE tokenizer imposes its own statistical "theory" of language onto a text, a theory biased towards common symbol collocations found in its training data. This pre-processing can effectively neutralize the very stylistic effects under study; a sequence of characters might be collapsed into a single token, artificially lowering its perplexity and rendering the author's stylistic choice invisible to the model. In computational studies of classical Chinese poetry, for instance, the fact that DeepSeek R1 represents *ai'ai* 暧暧 ("dimly") as four tokens [923, 247, 923, 247] but *yi'yi* 依依 ("softly") as one token [66660] can severely impact character-to-character comparisons (Kurzynski, Xu, and Feng 2025). To circumvent such inconsistencies, this study treats each Chinese character as a distinct token, allowing for a more fine-grained analysis of which characters in multi-character words and phrases elicit surprise. A custom character-level tokenizer was constructed by training on the first two million documents from the `Fineweb-Edu` corpus, and the resulting vocabulary was supplemented with single-character tokens from the widely used `bert-base-chinese` model (Devlin et al. 2019). The final vocabulary consists of 20,192 tokens and was used consistently across all experiments in this study.

Experiment 1: Macroscopic analysis of Maospeak

The first experiment is a macroscopic study examining how a language model's statistical worldview is altered when its linguistic environment becomes saturated with a powerful idiolect. "Maospeak" (*Mao yu* 毛语 or *Mao wenti* 毛文体 in its written form) refers to the militant, ideologically charged language style that emerged during the Mao Zedong era (1949–1976) in the People's Republic of China (Li 1998). While Mao himself was a creative rhetorician, the institutionalization of his discourse transformed his words into instruments of mass indoctrination, producing a ritualistic culture of near-compulsory recitation (Lifton 1956; Leese 2011; Kiely 2014). This engineered form of discourse became a central feature of everyday life, particularly during the Cultural Revolution (1966–1976), when proficiency in Maospeak became essential for survival amid massive outbursts of violence (Ji 2004; Juntao and Brady 2012; Schoenhals 2007). Though Mao passed away in 1976, his distinctive language continues to influence contemporary Chinese speech, from jingoistic discourse to colloquialisms like *xiaomie* (消灭, lit. "to annihilate") for finishing a meal or the encouraging phrase *haohao xuexi, tiantian xiang shang* (好好学习, 天天向上, "study well, every day upwards") (Link 2012, 2013; Barmé 1996).

The experiment tests a dual hypothesis. First, I hypothesize that a dominant political style has a representative phraseology that, through frequent exposure, becomes automatic in the Shklovskian sense. The experiment aims to identify these core expressions by tracking changes in model's loss, predicting that as the model is exposed to Maospeak, its loss will decrease significantly on ideologically reinforced phrases, specifically those that were either absent from the training data (Base Corpus) and/or frequently repeated in the fine-tuning data (Mao Corpus). Second, I hypothesize that as the model overfits on Maospeak, the stylistic and lexical patterns of literary prose (Novels Corpus) will become increasingly

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

Table 1. Corpus and model training parameters.

Parameter	Value
Corpus sizing	
Fineweb-Edu V. 2.1/4_5	17,790,513 documents, 58,267,562,012 characters
Base Corpus (filtered *Fineweb-Edu V. 2.1/4_5*)	1,000,000 chunks, 901,413,780 characters
Mao Corpus (*Selected Works of Mao Zedong*)	2,476 chunks, 1,770,917 characters
Novels Corpus (*Top 100 Chinese Novels*)	8,453 chunks, 8,619,170 characters
Model architecture (GPT–2 style)	
Transformer layers (*n*-layer)	16
Attention heads (*n*-head)	16
Embedding dimension (*n*-embd)	1,024
Trainable parameters	223,266,816
Vocabulary size	20,192
Pre-training hyperparameters	
Learning rate	2e–4 → 5e–5
Batch size and sequence length	32; 1,024
Epochs	1
Fine-tuning hyperparameters	
Learning rate	5e–5 → 1e–6
Batch size and sequence length	8; 1,024
Epochs	5

alien. The literary language, in essence, will become more difficult for the model to predict as its perceptual baseline shifts.

The experiment uses a GPT-2 style Transformer model pre-trained from scratch on the Base Corpus for one epoch and then fine-tuned for five epochs on the *Selected Works* (Table 1). The core analysis compares the model's performance on the Mao Corpus before fine-tuning ("Pre-trained") and after ("Epoch 5"). To track how the model internalizes this specific discourse, I deliberately deviate from standard practice and compute loss on the fine-tuning corpus itself. If $Q(x_i|x_{<i})$ denotes the probability that model Q assigns to token x_i given the preceding tokens, and if, following Shannon, the amount of surprise can be represented as a negative logarithm of the event's probability (unlikely events are surprising), then for a sequence of n tokens, the cross-entropy loss is the average of all individual surprises:

$$L = -\frac{1}{n}\sum_{i=1}^{n}\log Q(x_i|x_{<i})\,. \tag{1}$$

To identify the most significantly learned phrases, loss on each character was calculated across all documents in the Mao Corpus using a sliding window with a 256-character context. Only 16-character phrases (called "patterns" below) where every character was predicted with a full context were considered, eliminating bias from variable context lengths. For each valid pattern, aggregate loss was computed as the arithmetic mean of its character-level losses. The improvement on each pattern was quantified as the difference between the losses of the pre-trained and the fine-tuned model. All discovered patterns were then sorted in descending order of this loss drop.

As Tables 2 and 3 show, the most significant gains in predictive certainty occurred on phrases central to the era's political machinery. The model learned to anticipate the formulaic language used to categorize class adversaries, from the canonical list of "landlords, rich peasants, counter-revolutionaries, bad elements, and Rightists" (Rank 2) to the ubiquitous label for internal party targets: the "unrepentant capitalist-roader" (Rank 1) and the "small handful of capitalist-roaders in authority" (Rank 12). This familiarization extended to highly specific political targets, including internal factions like the "Gao Gang-Rao Shushi anti-Party alliance" (Rank 4), external ideological foes such as the "Soviet revisionist renegade clique" (Rank 3), and personalized lists of enemies, both domestic ("Zhang Bojun, Luo Longji, Zhang Naiqi," Rank 9) and international ("Chairman Chiang, Syngman Rhee, and Ngo Dinh Diem," Rank 11). The model also mastered the rigid, legalistic language of policy, such as the principle that the "chief culprits shall be punished, those who were coerced are not prosecuted, and those who perform meritorious service shall be rewarded" (Rank 16) and the formulation to "resolutely, thoroughly, wholly, and completely annihilate" the enemy (Rank 8). Beyond political labeling, the model also captured the discourse's distinctive rhetorical features, such as Mao's characteristic use of classical allusions repurposed for political ends, like the phrase "Sun Wukong drilling into Princess Iron Fan's belly to make trouble" (Rank 10), which co-opts a well-known episode from the classic Chinese novel *Journey to the West* 西游记 to describe subversive tactics.

These findings illustrate a process of Shklovskian familiarization, where a powerful idiolect turns slogans, labels, rhetorical strategies and even conversational tics into the automatic discursive background, a process which "replaces things with symbols" (Shklovsky 2017, 79). The new symbolic order functioned as political common sense, establishing the terms of legitimacy and effectively rendering the new regime's authority self-evident. This principle found its ultimate expression in Lin Biao's declaration

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

Table 2. Top 20 most significantly learned 16-character phrases with context.

Rank	Loss drop	Original phrase in context	English translation in context
1	4.1538	. . . 到底是自力更生快, 还是搞洋奴哲学快, 不是很清楚吗? 党内**那个不肯改悔的走资派口口声声说要**把国民经济搞上去, 却偏偏闭住眼睛不看事实 . . .	. . . Is it faster to rely on self-reliance, or is it faster to engage in the slavish comprador philosophy, isn't it clear? **That unrepentant capitalist-roader within the Party pays lip service to wanting** to get the national economy going, yet insists on closing his eyes to the facts. . .
2	4.1299	. . . 我们是站在占总人口的百分之九十五以上的人民大众一边, 绝不站在占总人口百分之四、**五的地、富、反、坏、右那一边**。在国际范围内也是这样 . . .	. . . We stand on the side of the great masses of the people, who make up more than ninety-five percent of the total population, and will never stand on **the side of the four or five percent–that of the landlords, rich peasants, counter-revolutionaries, bad elements, and Rightists.** It is the same on an international scale. . .
3	3.9319	. . . 在反对帝国主义和各国反动派的斗争中, 在**反对以苏修叛徒集团为中心的现代修**正主义的斗争中, 继承、捍卫和发展了马克思列宁主义 . . .	. . . in the struggle against imperialism and reactionaries of all countries, and in the struggle **against modern revisionism centered on the Soviet revisionist renegade clique**, he inherited, defended, and developed Marxism-Leninism. . .
4	3.9259	. . . 必须对那些钻不进去的人、浮在皮面上的人进行教育, 使他们都成为内行。**反对高岗、饶漱石反党联盟的斗争,** 将促使我们党大进一步 . . .	. . . it is necessary to educate those who cannot delve deep, who remain on the surface, so that they all become experts. **The struggle against the Gao Gang–Rao Shushi anti-Party alliance** will push our Party to make great strides forward. . .
5	3.7240	. . . 才有可能承认 "三民主义为中国**今日之必需, 本党愿为其彻底实现而**奋斗", 否则就没有这种可能了。这是共产主义和三民主义在民主革命阶段上的统一战线 . . .	. . . only then is it possible to recognize that "the Three People's Principles **are a necessity for China today, and our Party is willing to strive for their thorough realization**"; otherwise, such a possibility would not exist. This is the united front of Communism and the Three People's Principles in the stage of the democratic revolution. . .
6	3.6606	. . . 贫农领袖中, 从前虽有些确是有缺点的, 但是现在多数都变好了。他们自己在那里努力**禁牌赌, 清盗匪**。**农会势盛地方, 牌**赌禁绝, 盗匪潜踪 . . .	. . . Among the poor peasant leaders, although some indeed had shortcomings in the past, most have now improved. They themselves are working hard to **ban gambling and clear out bandits. In places where the peasant associations are powerful, gambling** is prohibited and bandits have vanished. . .
7	3.5018	.. 毛泽东同志关于开展无产阶级文革大革命的理论, 关于人民拥有**大鸣大放大辩论大字报"四大"权利**的理论, 关于对修正主义反动派造反有理的理论 . . .	. . . Comrade Mao Zedong's theory on launching the Great Proletarian Cultural Revolution, the theory that the people possess the **"Four Greats" rights of speaking out freely, airing views fully, holding great debates, and writing big-character posters**, the theory that it is right to rebel against revisionist reactionaries. . .
8	3.4810	. . . 就可以保证, 不但不会被敌人打散, 相反地, 一定要把日本侵略者及其忠**实走狗坚决、彻底、干净、全部地消**灭掉, 并且在消灭他们之后, 把一个新民主主义的中国建设起来 . . .	. . . it can be guaranteed that not only will we not be scattered by the enemy, but on the contrary, we must **resolutely, thoroughly, wholly, and completely annihilate** the Japanese aggressors and their faithful **running dogs**, and after annihilating them, build a new-democratic China. . .
9	3.4677	. . . 民主党派现在作风问题在其次, 主要是走那条路线的问题 . . . 是走章**伯钧、罗隆基、章乃器、陈仁炳、彭**文应、陆治、孙大雨那种反革命路线, 还是走什么路线?	. . . For the democratic parties now, the issue of work style is secondary; the main issue is which line to follow. Is it the counter-revolutionary line of people like **Zhang Bojun, Luo Longji, Zhang Naiqi, Chen Renbing, Peng** Wenying, Lu Zhi, and Sun Dayu, or some other line?
10	3.4510	. . . 就要准备一副清醒的头脑去对付对方采**用孙行者钻进铁扇公主肚子里兴妖作**怪的政策。只要我们精神上有了充分的准备, 我们就可以战胜任何兴妖作怪的孙行者 . . .	. . . one must be prepared with a clear head to deal with the other side's policy of **adopting Sun Wukong's tactic of drilling into Princess Iron Fan's belly to make trouble**. As long as we are fully prepared spiritually, we can defeat any trouble-making Sun Wukong. . .

that "one sentence of Chairman Mao's is worth ten thousand of ours," creating a high-stakes environment where a mere slip of the tongue could have fatal consequences (Schoenhals 1992, 19–20). The complementary phenomenon of defamiliarization can be illustrated by comparing the model's average perplexity on 16-grams and entire documents in both test corpora as it is progressively fine-tuned on Maospeak. As shown in Figure 1, a large difference exists even for the pre-trained model, suggesting that the stylistic gap between the Novels Corpus and the Base Corpus is wider than that between the Mao Corpus and the Base Corpus. As the model specializes in Mao's text, the more varied, lexically diverse, and structurally complex language of fiction becomes increasingly alien and harder to predict. This trade-off exemplifies what one might call "cognitive overfitting," a process where becoming an "expert" in a narrow linguistic domain, without access to external stimuli, leads to a quantifiable loss of fluency in another.

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

Table 3. Top 20 most significantly learned 16-character phrases with context (continued).

Rank	Loss drop	Original Phrase in Context	English translation in context
11	3.3185	. . . 现在杜勒斯就干这一套, 他就专扶什么蒋委员长、李承晚、吴庭艳这类人。我看是这样, 双方都怕, 但是他们怕我们比较多一点, 因此战争是打不起来的 . . .	. . . Now Dulles is up to this trick, he exclusively **supports people like what's-his-name Chairman Chiang, Syngman Rhee, and Ngo Dinh Diem**. I see it this way: both sides are afraid, but they are a bit more afraid of us, so a war is unlikely to break out. . .
12	3.3103	. . . 应迅速返回本地区、本单位进行斗批改, 把本单位被一**小撮走资本主义道路当权派篡夺的**权夺回来, 不要逗留在北京和其他地方 . . .	. . . should quickly return to their own regions and units to carry out struggle-criticism-transformation, to seize back the power in their units that was **usurped by the small handful of capitalist-roaders in authority**, and not to linger in Beijing or other places. . .
13	3.2756	. . . 他们用绳子捆绑了劣绅, 给他戴上高帽子, 牵着游**乡 (湘潭、湘乡叫游团, 醴陵叫游垅**)。他们那粗重无情的斥责声, 每天都有些送进绅士们的耳朵里去 . . .	. . . They tied up the evil gentry with ropes, put high hats on them, and led them while **parading through the villages (in Xiangtan and Xiangxiang it is called "parading through the groups," in Liling "parading through the ridges")**. The sound of their coarse and merciless rebukes entered the ears of the gentry every day. . .
14	3.2717	. . . 聂元梓: 王、关、戚插了一手。毛泽东: 你们反王、关、戚好嘛。你们搞串连**, 我也禁止不了**。**韩爱晶、蒯大富你**们不是好朋友吗? 你们两个以后还要作好朋友 . . .	. . . Nie Yuanzi: Wang, Guan, and Qi had a hand in it. Mao Zedong: It's good that you oppose Wang, Guan, and Qi. You engage in networking, **and I can't forbid it. Han Aijing, Kuai Dafu, you** two are good friends, aren't you? You two should remain good friends in the future. . .
15	3.1739	. . . 八、文艺上反对外国修正主义的斗争, 不能只捉丘赫拉**依之类小人物**。**要捉大的, 捉肖洛霍**夫, 要敢于碰他。他是修正主义文艺的鼻祖 . . .	. . . Eighth, in the struggle against foreign revisionism in literature and art, we cannot just catch **small figures like Chukhrai. We must catch the big ones, catch Sholokhov**; we must dare to touch him. He is the originator of revisionist literature and art. . .
16	3.1462	. . . 本军对于蒋方人员, 并不一概排斥, 而是采取分别对待的方针。这就是首**恶者必办, 胁从者不问, 立功者受奖**。对于罪大恶极的内战祸首蒋介石 . . .	. . . Our army does not reject all of Chiang's personnel, but adopts a policy of differential treatment. This is: **the chief culprits shall be punished, those who were coerced are not prosecuted, and those who perform meritorious service shall be rewarded**. As for the arch-criminal of the civil war, Chiang Kai-shek. . .
17	3.1218	. . . **《** 中共中央关于控制我国人口增长**问题致全体共产党员共青团员的公开信 》** 说 "有些同志担心, 一对夫妇只生育一个孩子, 将来会出现一些新的问题: 例如人口的 . . .	. . . The "Open Letter from the Central Committee of the Communist Party of China to All Communist Party and Communist Youth League Members on the **Problem of Controlling Our Country's Population Growth**" states, "Some comrades worry that if a couple has only one child, new problems will arise in the future: for example, the population's. . .
18	3.0652	. . . 全党全民的监督, 中央和地方的大多数干部的政治水平, 比**较一九五三年高、饶事件时期大为提**高了, 懂事了。庐山会议上这一场成功的斗争, 不就是证据吗?	. . . The supervision of the entire Party and people, and the political level of the majority of cadres at the central and local levels, have been **greatly raised compared to the period of the Gao–Rao Incident in 1953**; they have become more sensible. Isn't the successful struggle at the Lushan Conference proof of this?
19	3.0518	. . . 它的职权是: 召集全国人民代表大会会议, 解释法律, 制定法令, 派**遣和召回驻外全权代表, 接受外国使**节, 批准和废除同外国缔结的条约 . . .	. . . Its powers and functions are: to convene the session of the National People's Congress, to interpret laws, to enact decrees, to **appoint and recall plenipotentiary representatives to foreign states, to receive foreign envoys**, and to ratify and abrogate treaties concluded with foreign states. . .
20	3.0316	. . . 我们两国的社会条件有所不同, 但是目的是一致的。**毛:(面向贡代 · 塞杜) 你是搞教**育的。犯了罪的人也要教育。动物也可以教育嘛!	. . . The social conditions in our two countries are different, but our goal is the same. **Mao: (turning to Gontier-Sédouy) You are in education**. People who have committed crimes must also be educated. Even animals can be educated!

Experiment 2: Microscopic analysis

Transitioning from a macroscopic to a microscopic view, the second experiment examines how statistical familiarization manifests at the character level. To this end, I use perplexity, a more interpretable metric calculated by exponentiating the loss defined in Equation (1). A perplexity of 5, for instance, indicates that the model is as uncertain as if it were choosing the next token from five equally likely options. By plotting per-character perplexity for representative passages, we can observe precisely which linguistic elements are initially surprising and how fine-tuning reduces this surprise, effectively "indoctrinating" the model. The figures below illustrate this process, with the yellow line representing the base model's initial perplexity and the darkening shades of orange and red showing the progressive decrease across five epochs.

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

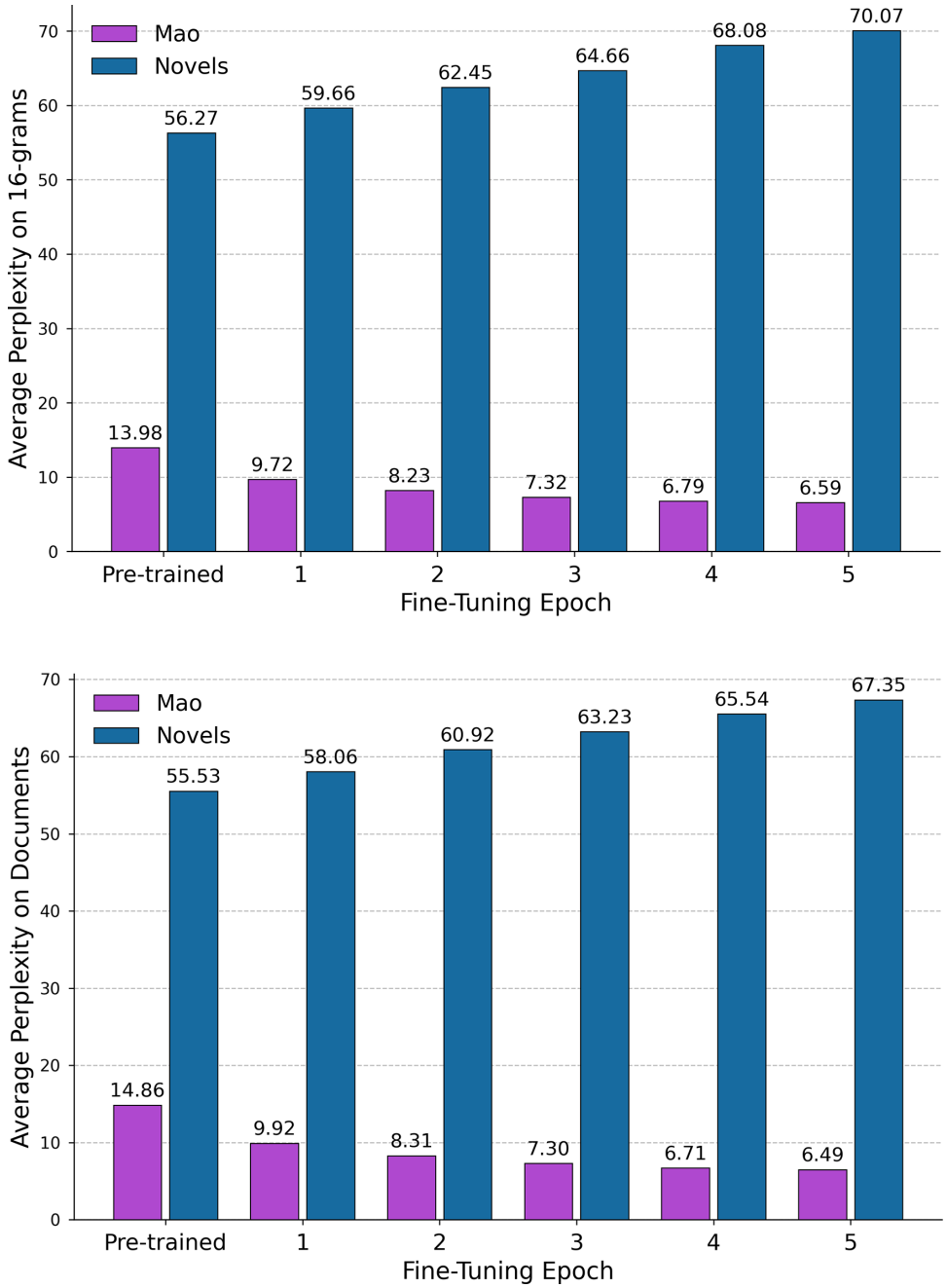


Figure 1. Comparison of average perplexities on 16-grams (top) and 1,024-token documents (bottom), as calculated on 1,000 trials with 10,000 samples per trial.

All plots exhibit a characteristic "waterfall" pattern: the first character of a key phrase surprises the model, but once this character is revealed, it strongly constrains what can follow, causing subsequent characters to cascade downward with near-certainty. This phenomenon, where the beginning of a sequence is more informative and thus harder to predict than its end, is a general feature of language modeling, whether applied to words (Pimentel, Cotterell, and Roark 2021; Pagnoni et al. 2025) or chapter breaks (Sun, Thai, and Iyyer 2022). What is distinctive about Maospeak is the length and frequency of such rigid sequences, which severely limit the range of valid continuations.

Bearing resemblance to what Shklovsky described as an automatized perception in which objects are recognized "by their initial features" (Shklovsky 2017, 79), this characteristic is clearly visible across the three examples (Figure 2). The first passage (top figure) highlights the model's internalization of characteristic rhetorical structures, featuring the militant formulation 坚决、彻底、干净、全部地消灭掉 ("resolutely, thoroughly, wholly, and completely annihilate," positions 26–39). The base model is initially very surprised by the sequence of four parallel adverbs. While 坚决 (resolutely) is a common word, the piling on of three more in this escalating sequence is a stylistic hallmark of Maoist rhetoric. The adverbs are followed by an equally characteristic term, *niugui sheshen* 牛鬼蛇神 (lit. "cow demons and snake spirits"), originally rooted in Buddhist sutras but repurposed by Mao for political agitation and persecution of intellectuals (Lee 2014, 49–51). An additional check reveals that this exact rhetorical structure is

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

used widely across the *Selected Works*, often in nearly identical contexts:

- 对于任何敢于反抗的反动派, 必须坚决、彻底、干净、全部地歼灭之 ("As for any reactionaries who dare to resist, they must be resolutely, thoroughly, wholly, and completely annihilated").
- 在此种情况下, 我们命令你们: 奋勇前进, 坚决、彻底、干净、全部地歼灭中国境内一切敢于抵抗的国民党反动派 ("In these circumstances, we order you: advance bravely, and resolutely, thoroughly, wholly, and completely annihilate all Kuomintang reactionaries within China who dare to resist").
- 一定要把日本侵略者及其忠实走狗坚决、彻底、干净、全部地消灭掉 ("The Japanese aggressors and their loyal lackeys must be resolutely, thoroughly, wholly, and completely annihilated").
- 也罢, 任何这类的欺骗, 必须坚决、彻底、干净、全部地攻击掉, 决不容许保留其一丝一毫 ("So be it; any such deception must be resolutely, thoroughly, wholly, and completely annihilated, with not the slightest trace allowed to remain").

The second passage (middle figure) showcases the learning of a canonical list of political labels: 地、富、反、坏、右 ("landlords, rich peasants, counter-revolutionaries, bad elements, and Rightists," positions 33–39). For the base model, this string of single characters is nonsensical and generates extremely high perplexity. Although the characters are common individually, their specific concatenation as an enumeration of class enemies is unique to Maoism.

The third passage (bottom figure) provides a clear example of the model learning specific ideological jargon, containing the phrase 在反对以苏修叛徒集团为中心的现代修正主义的斗争中 ("in the struggle against modern revisionism centered on the Soviet revisionist renegade clique"). The base model (yellow line) displays massive perplexity spikes on the key term 苏修叛徒集团 ("Soviet revisionist renegade clique," positions 40–44). This is a highly specific piece of Cold War-era verbiage that is rare in general-purpose text and thus highly surprising. Like in the previous example, with each epoch of fine-tuning, the perplexity on this exact phrase plummets as the model learns to recognize it as a core component of the discourse (Figure 2).

As Ji Fengyuan argues, linguistic engineering was central to Maospeak as a tool of political control (Ji 2004). Aligning with a weak interpretation of the Sapir–Whorf hypothesis, she suggests that language influences thought by making certain cognitive pathways more accessible. The constant repetition of slogans and labels activated associative networks and cognitive schemas tied to concepts like "class struggle" and "proletarian revolution," aligning public sentiment with ideological goals and providing ready-made justifications for denunciation and violence, effectively automating cruelty. However, Ji also cautions against an Orwellian view of language as a totalizing instrument of domination (Ji 2004, 134–143). The low-perplexity canyons are interspersed with less predictable tokens, revealing moments of instability; some grammatical connectives or less common phrasings elicit moderate surprise even after fine-tuning, representing points where Maospeak gives way to the more varied patterns of general language. Conversely, common function words like 的 (of), 在 (at) and 中 (in) consistently show low perplexity from the outset. The initial perplexity spikes are therefore reserved for elements stylistically and ideologically unique to Maospeak, underscoring the method's ability to isolate the process of familiarization with the specific features of a discourse against the backdrop of general linguistic competence.

Experiment 3: *N*-gram entropy

The previous experiments suggested that Maospeak operates by relying on a constrained set of high-frequency, low-perplexity phrases, making it more statistically compressible than other forms of discourse. To test this hypothesis quantitatively, I conducted an *n*-gram entropy analysis. The experiment compared the Mao Corpus against the Novels Corpus by repeatedly sampling 100,000 *n*-grams (from $n = 2$ to $n = 16$) from each and calculating the Shannon entropy. As a measure of predictability, lower entropy signifies a more repetitive text (approaching zero for a text with no variation), while higher entropy indicates greater linguistic variety (maximized in a uniform distribution of *n*-grams).

The results (Figure 3) support the initial hypothesis. At every *n*-gram size tested, the Mao Corpus exhibited lower entropy than the literary corpus, indicating a more constrained and formulaic linguistic system. The difference was most pronounced for short phrases; for bigrams ($n = 2$), the entropy of the Mao Corpus was 13.442 bits compared to 14.331 for the novels. This entropy gap narrowed as *n*-gram size increased, which is expected: as *n* grows, *n*-grams become so specific that they are unlikely to repeat within a fixed sample, causing the entropy of both corpora to approach the theoretical maximum for a sample of this size ($\log_2(100{,}000) \approx 16.6$ bits). The key finding is that the Mao Corpus remains measurably less entropic at smaller *n*-gram sizes, confirming that its repetitiveness is most evident in its foundational vocabulary and short collocations. The heavy repetition of specific *n*-grams, resulting in lower entropy, corresponds to the "waterfall" patterns observed in the fine-tuned model's perplexity scores.

Experiment 4: Impact on literature

Finally, I turn to works of literature to examine how the opposing impulses of familiarization and defamiliarization manifest in novels. By examining how three distinct modern Chinese authors engage with, repurpose, or resist predictable linguistic patterns, we can observe how literary art creates meaning through surprise. While familiarization reduces perplexity through repeated exposure, defamiliarization emerges (indirectly) as a by-product of overfitting and (directly) as an aesthetic intervention. Therefore, to explore this phenomenon, the following passages were selected subjectively for their aesthetic and stylistic significance.

The first passage (Figure 4, Top) comes from a pivotal scene in Chapter 24 of Zhang Wei's 张炜 (b. 1956) novel *The Ancient Ship* 古船 (1987). The story is set in Wali, an imagined microcosm of Chinese society which witnesses the vicissitudes of the communist revolution. In the passage, a "red-faced young man," representing the fervent revolutionary youth, delivers an incendiary speech at a public struggle session, aiming to mobilize the crowd to overthrow the town's authorities.

Many people recalled the hunger of those years, remembered the waves of devastation, and felt immense anger. Everyone shouted, "It is right to rebel! Down with the die-hard capitalist-roaders!" "If the enemy refuses to surrender, let them perish!" After the slogans, the young man continued his speech: "Those who are not afraid of being cut to a thousand pieces dare to pull the emperor off his horse. Revolutionary comrades, we are ready to shed our blood to make the whole world red. Comrades, let us unite and fight! Fight!"

不少人想起了那些年的饥饿、想起了一场场蹂躏, 无比愤怒。大家高喊:"造反有理! 打倒死不改悔的走资派!" "敌人不投降, 就让他灭亡!" 口号毕, 小伙子又继续演说:"舍得一身剐,

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

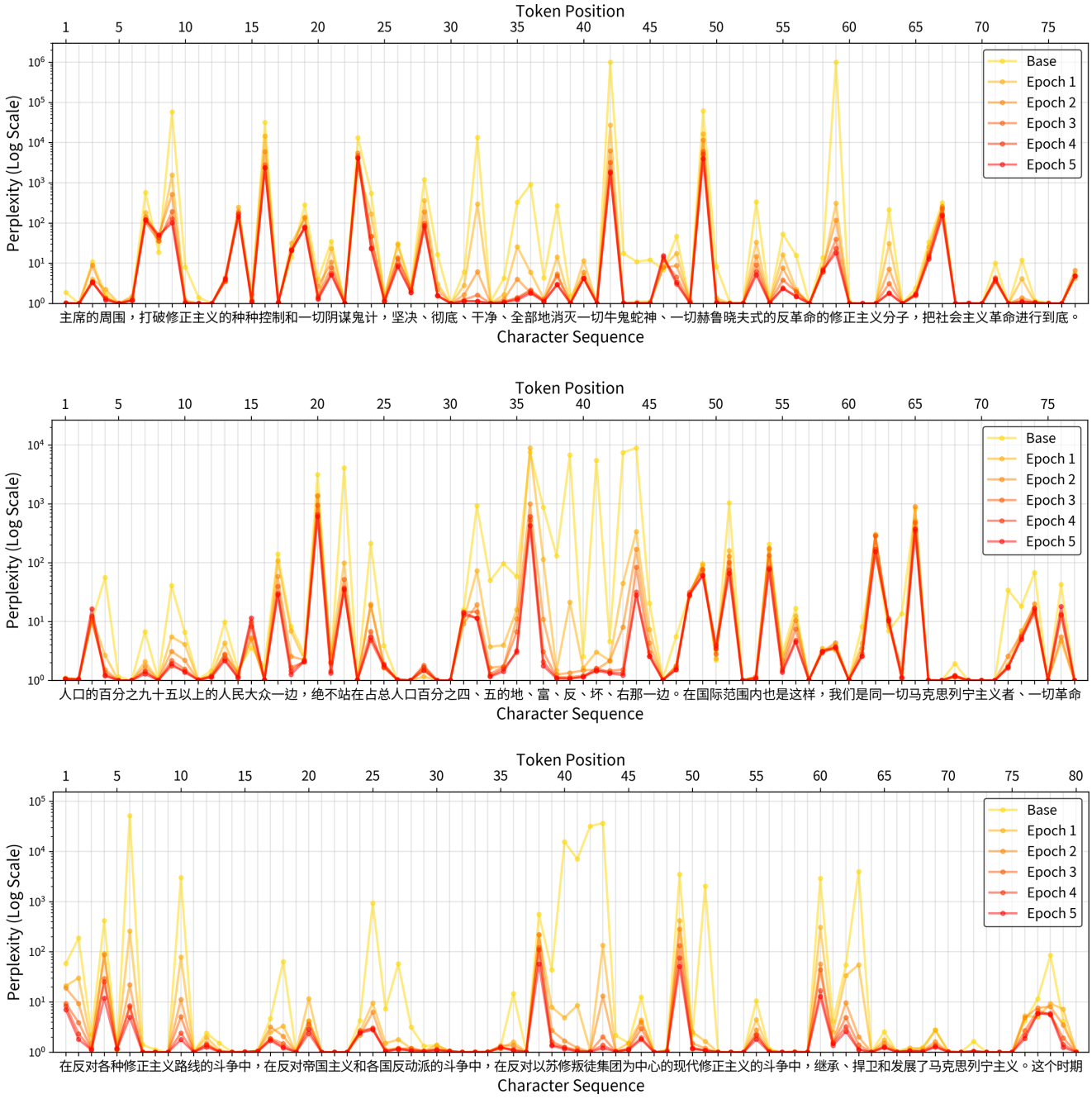


Figure 2. Per-character perplexity plots for three excerpts from the *Selected Works of Mao Zedong*. The yellow line (Base) shows the perplexity of the pre-trained model, while the lines from orange to red show the decreasing perplexity over five epochs of fine-tuning on the Mao Corpus. The *x*-axis displays the Chinese character sequence, and the *y*-axis shows perplexity on a logarithmic scale.

敢把皇帝拉下马, 革命的战友们, 我们甘洒一腔血, 换得全球一片红; 战友们, 让我们团结起来, 战斗! 战斗!"

Zhang Wei immerses the reader in the linguistic reality of the turbulent era that he himself experienced as a child. The speech is a barrage of Maoist-era slogans showcasing the romanticization of violence ("spill our blood") and rhythmic chants ("Fight! Fight!"). The most potent phrase is the proverb, "Those who are not afraid of being cut to a thousand pieces dare to pull the emperor off his horse" (positions 29–41), which comes from another Chinese classic, the *Dream of the Red Chamber*, and which gained new life through Mao's use. By placing this slogan in the young man's mouth, Zhang Wei not only highlights the central paradox of the era: the nation's supreme leader encouraging violent rebellion against all "emperors," but also masterfully imitates the dominant language style of the period. The Maoist slogan is the only part of the sequence that experienced a significant drop in perplexity (Figure 4, Top). As such, the method was able to identify the Maospeak embedded within the novel.

The second case study examines a passage from *Farewell My Concubine* 霸王别姬, the seminal 1985 novel by Hong Kong author Lilian Lee 李碧华 (b. 1958). The novel chronicles the tumultuous relationship between two Peking opera performers, Duan Xiaolou and Cheng Dieyi, against the backdrop of China's political upheavals. The passage in question appears early in the novel, as the protagonists, then young boys, endure the brutal training of an opera troupe. After Xiaolou is caught secretly helping Dieyi

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

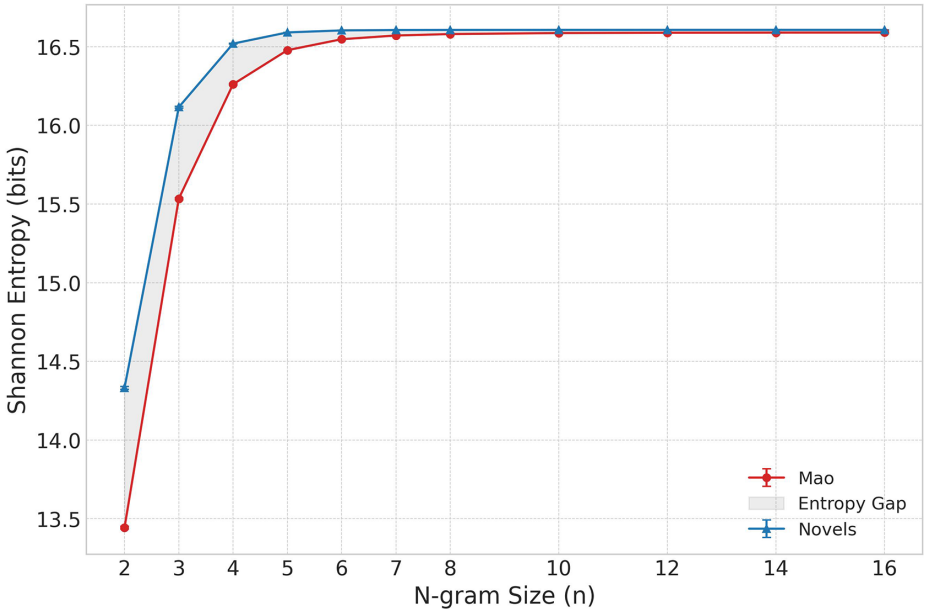


Figure 3. Shannon entropy comparison between Mao's *Selected Works* and Chinese novels across different *n*-gram sizes on 1,000 trials.

ease the pain of a torturous leg-stretching exercise, Master Guan punishes the entire troupe with collective beating. The passage offers a piercing narrative insight into the master's psychology and the constrained nature of masculinity that defines the novel's world.

There are too many people whose lives are not as they wish. Women can cry, and children can cry, but a dignified man can only vent through different feigned excuses: a heroic sneeze, a ferocious yawn, a roar at the weak and small who cannot fight back. All this surging and churning, is it not because for a petty man, the chance to feel proud and elated is nowhere to be found?

不如意的人太多了, 女人可以哭, 孩子可以哭, 但堂堂男子, 只能假不同的借口抒泄: 轰烈地打喷嚏、凶狠地打哈欠、向无法还手的弱小吼叫。这些汹涌澎湃, 自是因为小丈夫, 吐气扬眉机会安在?

The perplexity distribution for this passage (Figure 4, Middle) shows how literary language creates its effects through moments of surprise. The highest perplexity spikes occur on the creative juxtapositions: 凶狠地打哈欠 ("a ferocious yawn," 45–51) or 向 … 弱小吼叫 ("to roar at the weak," 59–62). These phrases elicit high perplexity likely because figurative language is less probable than literal phrases (Pedinotti et al. 2021). Even more telling is the passage's use of pre-modern and classical Chinese phrasings, which become increasingly unfamiliar to the model as it is fine-tuned on Maospeak. For instance, the adjective 堂堂 in 堂堂男子 ("dignified man") is a formal, literary term for "imposing" or "grand" that ironically evokes here a traditional sense of masculinity, a register alien to the functional dialect of Maospeak. Likewise, the final question, 机会安在 (lit. "the chance is where?"), uses the character 安 in its classical function as an interrogative pronoun ("where"), a usage largely replaced by words like 哪里 in modern vernacular and mostly absent from the Mao Corpus. The fine-tuned model finds both of these literary phrasings even more anomalous than it did initially, causing their perplexity to increase.

These high-entropy word choices, which slow down the narrative, are set against a backdrop of linguistic convention (Bizzoni, Feldkamp, and Nielbo 2024) and narrative scaffolding that glues together the surprising tokens. The passage opens with just such a phrase: 不如意的人太多了 ("too many people whose lives are not as they wish"). This is a low-perplexity, almost proverbial, statement that eases the reader into the scene by establishing a familiar emotional context. This kind of predictable framing anchors the unique description of events in familiar literary conventions, signaling the narrative's fictional nature and making the subsequent, high-perplexity surprises more effective.

The final case study focuses on Hong Kong writer Dung Kai-cheung 董启章 (b. 1967), known for his postmodern narratives merging memory, technology, and urban space. A distinctive feature of Dung's style is his interspersion of standard written Chinese with Cantonese, the non-normative language spoken in Guangdong, Hong Kong, and numerous overseas Chinese communities in Southeast Asia and North America. His novel *Works and Creations: As Vivid As Real* 天工开物 · 栩栩如真 (2005) interweaves fictional and autobiographical storylines. In the autobiographical part, Dung's grandmother, Lung Gam Juk, confides in her husband a painful memory of separation from her brother, which is when she first realized her extraordinary ability to hear radio waves:

The next day, Lung Gam Juk began packing her luggage. That was the first time I heard the sound in the air. Lung Gam Juk told Dung Fu. The upright Dung Fu nodded. He studied electronics and didn't believe in esoteric or supernatural things, but he didn't argue or question his newly-wed wife's words.

第二天起来, 龙金玉就开始收拾她的行李。嗰次系我第一次听到空中既声音。龙金玉告诉董富。正直人董富点点头。他是学科技的, 他不相信神秘和超自然的事物, 但他没有反驳或质疑新婚妻子的说法。

This seemingly straightforward scene blends free indirect speech, unmarked direct quotations, and internal monologue. The

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

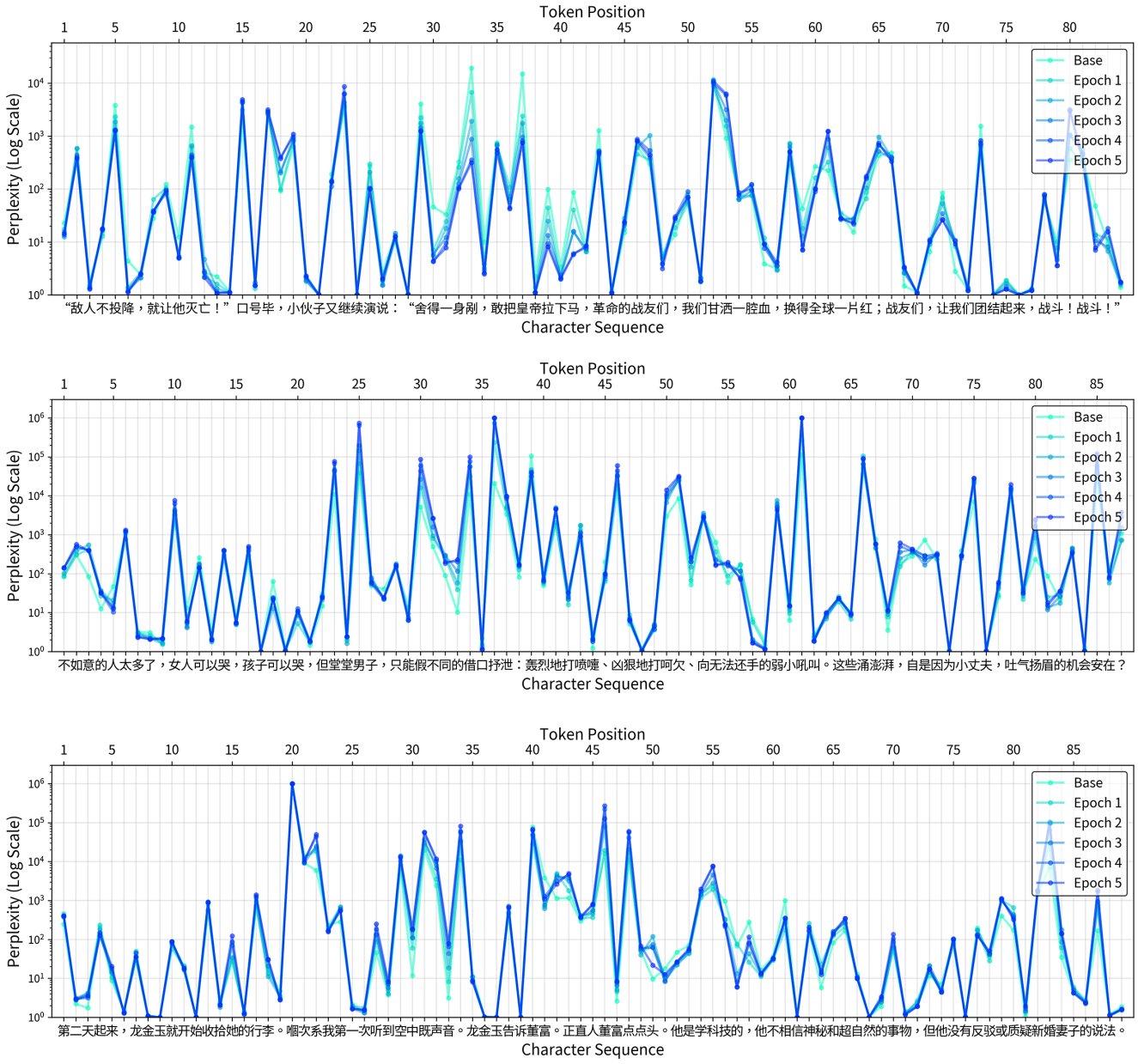


Figure 4. Per-character perplexity plots for three excerpts from novels. The light blue line (Base) shows the perplexity of the pre-trained model, while the darker lines show the decreasing perplexity over five epochs of fine-tuning on the Mao Corpus. The *x*-axis displays the Chinese character sequence, and the *y*-axis shows perplexity on a logarithmic scale. *Top*: Zhang Wei's *The Ancient Ship*; *Middle*: Lilian Lee's *Farewell My Concubine*; *Bottom*: Dung Kai-cheung's *Works and Creations*.

perplexity plot (Figure 4, Bottom) shows how Cantonese terms like 嗰 (“that”), 系 (“to be”) and 既 (“of”) disrupt the perplexity arc. The surprise is enhanced because the Cantonese statement is not announced with quotation marks, making an abrupt switch from the narrator's to the character's voice. Meanwhile, novel-specific terms like “upright” (正直人) and character names challenge the model's predictive capacities, providing a clear example of what empirical literary studies call “foregrounding” (Peer et al. 2021). As such, the heteroglossic novel “puts the brakes on perception” (Shklovsky 2017, 94), challenging both human and machine readers. Written Cantonese contains many characters that are absent from standard written Chinese but are essential to transcribe words used by Cantonese speakers on a daily basis (Snow 2004). Many Hong Kong writers use these non-standard characters to signal the cultural distinctiveness of their work, and some go as far as to write entire novels in Cantonese. On a technical level, Cantonese expressions resist prediction because they are underrepresented in training data, highlighting the normative assumptions underlying most Chinese LLMs (Xiang et al. 2025; Jiang et al. 2025).

This use of a topolect as a defamiliarizing device is a classic literary strategy. For language to be poetic, Shklovsky argued after Aristotle, it ought to have “the character of the foreign” (Shklovsky 2017, 94). He observed that when a literary establishment exhausts the standard language, it turns to regional vernaculars, and so “folk language and literary language have changed places” (Shklovsky 2017, 95). Although Cantonese can boast significantly richer cultural heritage than “modern standard Chinese,” Dung's insertion of Cantonese speech operates through similar mechanisms. By deliberately disrupting the automatized perception of written Chinese,

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

the author forces readers—or at least those accustomed only to the standard tongue—to encounter the cultural specificity of his text.

Discussion

The dialectic of familiarity and surprise

Literary critics since Shklovsky have often focused on defamiliarization as the primary function of art, dismissing the familiar as "worn-out rhetoric" and "jaded conventions" (Resseguie 1991). This perspective ignores the fact that no element of language is inherently familiar or conventional; such qualities emerge only through repeated exposure and cultural reinforcement. The present study has thus approached style not as a static property of a text but as a dynamic, processual engagement between a reader and a text unfolding in time.

The stylometric analysis of Maospeak offers a case study in how familiarity might be achieved in language. The relentless reinforcement of its phraseology creates the "economy of mental effort" that Shklovsky critiqued and which, as cognitive stylistician Reuven Tsur explains, fosters a "psychological atmosphere of control, certainty, and patent purpose" (Tsur 2008, 20). This effect is achieved by radically collapsing the probabilistic space of expression around a few high-stakes binaries – class struggle, revolution versus counter-revolution, East versus West – making dissent not only politically dangerous but also lexically improbable. In a similar vein, the institutional promotion of a single "standard Chinese" marginalizes non-normative languages. Faced with such top-down regulation, cognitive studies in the humanities have focused on the bottom-up "violence against cognitive processes" (Tsur 2008, 4) – the ways literature strikes back and interferes with habituated perception.

Yet the above experiments also indicate that a singular focus on defamiliarization overlooks the crucial role played by the familiar in grounding perception. Recent years have witnessed a renewed interest in the "eventness" of narratives (Baldassano et al. 2017; Sims, Park, and Bamman 2019; Gius and Vauth 2022), but singular events are meaningful only within a sea of stability where "things do not happen all the time" (Sayeau 2018). This principle applies directly to perplexity: a sequence of completely random characters would certainly be unfamiliar and perplexing, but it would not defamiliarize. Wendy O'Brien suggests that, in addition to its defamiliarizing function, fiction can also "make the familiar and the ordinary familiar again" (O'Brien 2007). This important restorative function of fiction lies in its ability to re-establish meaning by arranging overwhelming events into a temporal sequence, allowing the reader to "adapt to her new history" and "tell the time" again (O'Brien 2007).

In this context, Patrick Colm Hogan's notion of "non-anomalous surprise" offers a compelling theoretical alternative that might account for the findings of this project (Hogan 2016, 20–36). Faced with the question whether aesthetic pleasure is a function of predictability or surprise, Hogan differentiates between focal and non-focal aspects of the aesthetic target: "focal aspects would be pleasurable to the degree that they foster non-anomalous surprise whereas non-focal aspects would be valued primarily for predictability" (Hogan 2016, 27). In Hogan's cognitive framework, which echoes the figure-ground hypothesis of Gestalt psychology, having a predictable part of a work would be valuable even if the main aesthetic pleasure comes from the unpredictability of another part. He speculates that in music, "aesthetic pleasure results from the predictability of the non-focally processed rhythmic cycles in combination with the non-anomalous surprise of the focally processed melody" (Hogan 2016, 27). We can observe a similar phenomenon in novelistic texts, where the familiar narrative rhythms are disrupted with high-perplexity tokens, leading to rapid oscillations in perplexity arcs. Although such tokens are surprising, they are not anomalous, because the reader is able to recognize their meaning and function in the sequence once they occur.

Redefining style as a cognitive signature

The tension between a predictable ground and a surprising figure also allows us to redefine style in cognitive and probabilistic terms and to move beyond the prevailing definition of style as something that affects only *how* information is conveyed. This model underpins computational style transfer (Gatys, Ecker, and Bethge 2016; Shen et al. 2017; Jin et al. 2022), which can render a family photo in van Gogh's style or rewrite a hamburger description in Du Fu's verse. Instead, my project supports the Formalist refusal of the form-content separation, suggesting that stylistic differences are better understood as "varying relationships between highly categorized and pre-categorial information" (Tsur 2008, 20). From this perspective, style is the signature of how a text manages historically specific readerly expectations by orchestrating the interplay between clear-cut figures and their low-differentiated backgrounds. A meeting of two characters may provide a schematic foil for a surprising dialogue; the irruption of Cantonese, as seen in Dung Kai-cheung's work, challenges the presumed naturalness of standard Chinese (Yeung 2023). Maospeak operates more insidiously, pushing highly categorized political abstractions into the non-focal spectrum until the ideological discourse becomes a subconscious, low-perplexity rhythm.

From an LLM perspective, this model of style is directly measurable. The "norm" is the set of regularities learned during training; any perplexity above 1 indicates a choice. To speak with minimum perplexity would require quoting a memorized dataset without variation, a scenario glimpsed in the institutionalized forms of Maospeak which elevated self-similarity to an unprecedented degree. Consider this excerpt from Mao's seminal speech "Serve the People":

> *To die for the People's interests is weightier than Mount Tai; to die in service of fascists, or for those who exploit and oppress the People, is lighter than a feather. Comrade Zhang Side died for the People's interests, and his death was weightier than Mount Tai.*
>
> 为人民利益而死, 就比泰山还重; 替法西斯卖力, 替剥削人民和压迫人民的人去死, 就比鸿毛还轻。张思德同志是为人民利益而死的, 他的死是比泰山还要重的。

Such rigid syllogisms reinforce a global discourse through local, fractal-like iterations (Drożdż et al. 2016). Literature, by contrast, is not a self-perpetuating dataset but a dynamic interplay of fuzzy quotations and their creative recontextualizations, which together generate the "different ways of de-automatizing things" (Shklovsky 2017, 81). While a recent study argues for the presence of fractality in Chinese prose (Liu et al. 2023), this finding is based on sentence length variability rather than semantics. Curiously, the authors do note that "Maoist-era novels of socialist realism depicting rural China are among the narratives with the highest Hurst values [indicating high fractality]" (Liu et al. 2023, 609). The present study has argued that literature resists the self-similar, low-perplexity patterns of engineered languages, but further research

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

is needed to explore this tension. In particular, we have noted that Maospeak itself draws heavily upon literary quotations; it is the cognitive effect of repetition that differentiates it from literature. Where literature recontextualizes a quotation to defamiliarize and provoke thought, Maospeak repeats it to familiarize and ultimately foreclose it.

Perplexity arcs and emotional arcs

Finally, the focus on the unfolding of historically grounded readerly experience aligns this project with the analysis of narrative "emotional arcs" (Teodorescu and Mohammad 2023), a concept pioneered by the novelist Kurt Vonnegut, who proposed that stories could be plotted on a simple graph of fortune against time, identifying fundamental shapes like the "Man in a Hole" archetype. This intuitive idea was later operationalized by a wave of digital humanities research, which applied quantitative methods like stylometry and lexicon-based sentiment analysis to large corpora of novels, computationally identifying a set of recurring emotional trajectories and character patterns (Burrows 1987; Jockers 2013; Reagan et al. 2016; Rebora 2023). However, this methodology has been the subject of critical refinement. For example, in her book *The Shapes of Stories*, Katherine Elkins argues that scholars must move beyond a one-size-fits-all approach, demonstrating that different sentiment analysis models are required to capture the unique "affective fingerprint" of each narrative (Elkins 2022). Elkins advocates for an ensemble, critic-in-the-loop methodology that prioritizes the singular patterns of individual texts over generalized shapes.

The "perplexity arc" proposed in this study operates on a parallel, yet distinct, cognitive axis. A moment of emotional intensity might be rendered in simple, low-perplexity language to create a blunt effect. Conversely, a seemingly neutral description could be delivered in highly complex, high-perplexity prose, signaling its importance. The present approach is not immune to the critiques leveled against its affective counterpart; a single focus on local, token-level surprise risks overlooking how meaning and affect are constructed over larger spans of text. Shklovsky's examples, for instance, operate mostly at the macro-level of plot and perception, such as Tolstoy's use of a horse as a narrator. In this article, I have suggested that such disruptions can be measured at the most fundamental level (the moment-to-moment encounter with language) and that such micro-level surprises are the building blocks of larger-scale formal effects. A crucial next step is to bridge these scales, moving from phrases to sentences and beyond, an accumulation of linguistic contexts that has shown parallels in both artificial and human cognition (Tikochinski et al. 2025). By integrating these multiple cognitive dimensions and analytical scales, future work can create a far richer, multi-layered model of the reading experience that maps the complex interplay between emotional valence, cognitive surprise and the hierarchical structuring of literary meaning.

Conclusion

More than a century after Shklovsky's "Art as Device" laid the groundwork for theories from New Criticism to narratology (Wake 2013), his focus on defamiliarization continues to inspire literary studies. As complementary perspectives have noted, the aesthetic role of "familiarization" is equally crucial (O'Brien 2007; Vos 2024). The present study has argued that it is in between these two impulses that the Formalist project joins hands with information theory. Each style possesses a distinct cognitive signature, a specific strategy for managing a reader's predictive processing by balancing familiar grounding with moments of perplexity.

Cognitive stylistics explores the link between textual structure and perceived effect (Tsur 2008), while computational stylometry measures textual patterns (Argamon, Burns, and Dubnov 2010). By integrating these fields through the predictive lens of LLMs, I propose "cognitive stylometry" as a new mode of literary inquiry that turns the common critique of language models as "glorified autocomplete" into a methodological advantage and a new viewpoint. In a world where "everything is predictable" (Chivers 2024), literature can still teach us something about freedom and choice.

Limitations

Several limitations of this study should be acknowledged. First, the results are contingent on the chosen GPT-2 style architecture and the `Fineweb-Edu` corpus, which serves as the "neutral" baseline of modern Chinese. A different, perhaps larger, model or a baseline corpus with a different composition (e.g., one richer in classical literature or web-based text) would inevitably alter the specific perplexity values and could shift the perceived boundary between the familiar and the unfamiliar. Furthermore, the fixed context window of 256 tokens imposes an artificial boundary on the model's predictive horizon, which cannot fully replicate the long-term memory and rich contextual understanding of a human reader. The very fact that each literary excerpt in this study required a brief prose introduction highlights this gap. While there is no scholarly consensus on the optimal context length for such comparisons (Fang et al. 2024), and internal tests with different window sizes confirmed the overall trends and identified patterns, this parameter remains a significant variable.

Second, it bears repeating that the study's core analogy, which equates perplexity with cognitive surprise, is a deliberate simplification. Human reading involves embodied cognition, emotional resonance and multi-modal knowledge that the LLMs employed in this study lack, just as fine-tuning only approximates the socio-historical saturation of Maospeak. New model architectures and human-centered artificial intelligence might reduce these gaps, allowing scholars to better model historical contexts and circumstances.

Finally, the selected hyperparameters, while supported by local testing, necessarily impact the final results. Notably the fine-tuning learning rate and the gradient norm clipping influence the learning process and, if not carefully adjusted, might lead to "catastrophic forgetting" of the base model. An alternative setting would be to progressively contaminate the training dataset and train a series of different models, thus bypassing the fine-tuning stage. This approach is less practical than the one adopted in this study, however, as fine-tuning is less computationally expensive than training entire models from scratch.

Acknowledgements. The author would like to express his gratitude to the anonymous reviewers for their helpful comments and questions. Special thanks are also extended to the organizers and participants of the NLP4DH 2023 conference.

Data availability statement. The corresponding code and training settings can be found in the online repository (https://github.com/mcjkurz/cogstyl).

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

Author contributions. Maciej Kurzynski — Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing.

Funding statement. None.

Disclosure of use of AI tools. Claude Opus 4 was used to create the figures included in this document and to generate a *Streamlit* application for easier perusal of identified patterns.

Competing interests. The author declares none.

Supplementary Material. To view supplementary material for this article, please visit http://doi.org/10.1017/chr.2025.10020.

References

Argamon, Shlomo, Kevin Burns, and Shlomo Dubnov, eds. 2010. *The Structure of Style: Algorithmic Approaches to Understanding Manner and Meaning*. Springer. https://doi.org/10.1007/978-3-642-12337-5.

Baldassano, Christopher, Janice Chen, Asieh Zadbood, et al. 2017. Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* 95 (3): 709–721.e5. https://doi.org/10.1016/j.neuron.2017.06.041.

Barmé, Geremie R. 1996. *Shades of Mao: The Posthumous Cult of the Great Leader*. Armonk, New York: M. E. Sharpe.

Bizzoni, Yuri, Pascale Feldkamp, and Kristoffer Laigaard Nielbo. 2024. Global Coherence, Local Uncertainty: Towards a Theoretical Framework for Assessing Literary Quality. In *Proceedings of the Computational Humanities Research Conference 2024. CEUR-WS.org*.

Bizzoni, Yuri, Tom S. Juzek, Cristina España-Bonet, et al. 2020. How Human is Machine Translationese? Comparing Human and Machine Translations of Text and Speech. In *Proceedings of the 17th International Conference on Spoken Language Translation*, 280–290. Association for Computational Linguistics. https://aclanthology.org/2020.iwslt-1.34/.

Bloch, Maurice, ed. 1975. *Political Language and Oratory in Traditional Society*. London: Academic Press.

Bohrn, Isabel, Ulrike Altmann, Oliver Lubrich, et al. 2012. Old Proverbs in New Skins: An fMRI Study on Defamiliarization. *Frontiers in Psychology* 3: 204. https://doi.org/10.3389/fpsyg.2012.00204.

Bonhage, Corinna E., Jutta L. Mueller, Angela D. Friederici, et al. 2015. Combined Eye Tracking and fMRI Reveals Neural Basis of Linguistic Predictions during Sentence Comprehension. *Cortex* 68: 33–47. https://doi.org/10.1016/j.cortex.2015.04.011.

Brown, Tom B., Benjamin Mann, Nick Ryder, et al. 2020. Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems (NeurIPS)* 33: 1877–1901. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

Burrows, John F. 1987. Word-Patterns and Story-Shapes: The Statistical Analysis of Narrative Style. *Literary and Linguistic Computing* 2 (2): 61–70. https://doi.org/10.1093/llc/2.2.61.

Caucheteux, Charlotte, and Jean-Rémi King. 2022. Brains and Algorithms Partially Converge in Natural Language Processing. *Communications Biology* 5 (134). https://doi.org/10.1038/s42003-022-03036-1.

Chakraborty, Megha, S.M Towhidul Islam Tonmoy, S M Mehedi Zaman, et al. 2023. Counter Turing Test (CT2): AI-Generated Text Detection is Not as Easy as You May Think – Introducing AI Detectability Index (ADI). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, edited by **Houda Bouamor, Juan Pino, and Kalika Bali**, 2206–2239. Association for Computational Linguistics, December. https://doi.org/10.18653/v1/2023.emnlpmain.136.

Chivers, Tom. 2024. *Everything Is Predictable: How Bayesian Statistics Explain Our World*. Atria/One Signal Publishers.

Chomsky, Noam. 2023. The False Promise of ChatGPT. *The New York Times* (March). https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html.

Clark, Andy. 2013. Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *Behavioral and Brain Sciences* 36 (3): 181–204.

Clark, Andy. 2016. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.

Dehaene, Stanislas, Laurent Cohen, Mariano Sigman, et al. 2005. The Neural Code for Written Words: A Proposal. *Trends in Cognitive Sciences* 9 (7): 335–341. https://doi.org/10.1016/j.tics.2005.05.004.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, et al. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies , Volume 1 (Long and Short Papers)*, 4171–4186. Association for Computational Linguistics, June. https://doi.org/10.18653/v1/N19-1423.

Drożdż, Stanisław, Paweł Oświęcimka, Andrzej Kulig, et al. 2016. Quantifying Origin and Character of Long-Range Correlations in Narrative Texts. *Information Sciences* 331: 32–44. https://doi.org/10.1016/j.ins.2015.10.023.

Elam, Michele. 2023. Poetry Will Not Optimize; or, What Is Literature to AI? *American Literature* 95 (2): 281–303. https://doi.org/10.1215/00029831-10575077.

Elkins, Katherine. 2022. *The Shapes of Stories: Sentiment Analysis for Narrative*. Cambridge University Press. https://doi.org/10.1017/9781009270403.

Enkvist, Nils Erik, John Spencer, and Michael Gregory. 1964. *Linguistics and Style*. Oxford University Press.

Epstein, Mikhail N. 1990. *Ideolanguage: An Inquiry into the Language of Soviet Ideology*. Occasional Paper 237. Washington, D.C.: The Kennan Institute for Advanced Russian Studies, The Wilson Center.

Evert, Stefan, Thomas Proisl, Frederike Jannidis, et al. 2017. Understanding and Explaining Delta Measures for Authorship Attribution. *Digital Scholarship in the Humanities* 32: 4–16. https://doi.org/10.1093/llc/fqx023.

Fang, Lizhe, Yifei Wang, Zhaoyang Liu, et al. 2024. What is Wrong with Perplexity for Long-context Language Modeling? *arXiv preprint arXiv:2410.23771* (October). https://doi.org/10.48550/arXiv.2410.23771.

Feghhi, Ebrahim, Nima Hadidi, Bryan Song, et al. 2024. What Are Large Language Models Mapping to in the Brain? A Case Against Over-Reliance on Brain Scores. *arXiv preprint arXiv:2406.01538*, https://doi.org/10.48550/arXiv.2406.01538.

Fodor, James, Carsten Murawski, and Shinsuke Suzuki. 2025. When Word Order Matters: Human Brains Represent Sentence Meaning Differently from Large Language Models. *bioRxiv*, https://doi.org/10.1101/2025.07.19.665701.

Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. 2016. Image Style Transfer Using Convolutional Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2414–2423. https://doi.org/10.1109/CVPR.2016.265.

Gius, Evelyn, and Michael Vauth. 2022. Towards an Event-Based Plot Model: A Computational Narratology Approach. *Journal of Computational Literary Studies* 1 (1). https://jcls.io/article/id/110/.

Henderson, John M., Wonil Choi, Matthew W. Lowder, et al. 2016. Language Structure in the Brain: A Fixation-Related fMRI Study of Syntactic Surprisal in Reading. *NeuroImage* 132: 293–300. https://doi.org/10.1016/j.neuroimage.2016.02.050.

Hendrycks, Dan, Mantas Mazeika, and Thomas Woodside. 2023. An Overview of Catastrophic AI Risks. *arXiv preprint arXiv:2306.12001*, https://arxiv.org/abs/2306.12001.

Herrmann, J. Berenike, Karina van Dalen-Oskam, and Christof Schöch. 2015. Revisiting Style, a Key Concept in Literary Studies. *Journal of Literary Theory* 9 (1): 25–52. https://doi.org/10.1515/jlt-2015-0003.

Hogan, Patrick Colm. 2016. *Beauty and Sublimity: A Cognitive Aesthetics of Literature and the Arts*. Cambridge University Press. https://doi.org/10.1017/CBO9781316335604.

Holtzman, Ari, Jan Buys, Li Du, et al. 2020. The Curious Case of Neural Text Degeneration. In *Proceedings of the International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1904.09751.

Hughes, Geoffrey. 2009. *Political Correctness: A History of Semantics and Culture*. Malden, MA: Wiley-Blackwell.

Jelinek, F., R. L. Mercer, L. R. Bahl, et al. 1977. Perplexity—A Measure of the Difficulty of Speech Recognition Tasks. *The Journal of the Acoustical Society of America* 62 (S1). https://doi.org/10.1121/1.2016299.

Ji, Fengyuan. 2004. *Linguistic Engineering: Language and Politics in Mao's China*. University of Hawai'i Press.

Jiang, Jiyue, Pengan Chen, Liheng Chen, et al. 2025. How Well Do LLMs Handle Cantonese? Benchmarking Cantonese Capabilities of Large Language Models. In *Findings of the Association for Computational Linguistics: NAACL 2025*, edited by **Luis Chiruzzo, Alan Ritter, and Lu Wang**, 4464–4505. Association for Computational Linguistics, April. https://doi.org/10.18653/v1/2025.findings-naacl.253.

Jin, Di, Zhijing Jin, Zhiting Hu, et al. 2022. Deep Learning for Text Style Transfer: A Survey. *Computational Linguistics* 48 (1): 155–205. https://doi.org/10.1162/coli_a_00426.

Jockers, Matthew L. 2013. *Macroanalysis: Digital methods and literary history*. University of Illinois Press.

Juntao, Wang, and Anne-Marie Brady. 2012. Sword and Pen: The Propaganda System of the People's Liberation Army. In *China's Thought Management*, edited by **Anne-Marie Brady**, 101–117. Routledge.

Kiely, Jan. 2014. *The Compelling Ideal: Thought Reform and the Prison in China, 1901-1956*. Yale University Press.

Kirchenbauer, John, Jonas Geiping, Yuxin Wen, et al. 2023. A Watermark for Large Language Models. In *Proceedings of the 40th International Conference on Machine Learning*, edited by **Andreas Krause, Emma Brunskill, Kyunghyun Cho**, et al., 202: 17061–17084. Proceedings of Machine Learning Research. PMLR, 23–29 Jul. https://proceedings.mlr.press/v202/kirchenbauer23a.html.

Kirschenbaum, Matthew. 2023. Prepare for the Textpocalypse. *The Atlantic* (March). Accessed July 31, 2025. https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writinglanguage-models/673318/.

Klemperer, Victor. 2013. *The Language of the Third Reich: LTI – Lingua Tertii Imperii*. Translated by **Martin Brady**. London: Bloomsbury Academic.

Kumar, Sreejan, Theodore R. Sumers, Takateru Yamakoshi, et al. 2024. Shared Functional Specialization in Transformer-Based Language Models and the Human Brain. *Nature Communications* 15 (5523). https://doi.org/10.1038/s41467-024-49173-5.

Kurzynski, Maciej. 2023. The Stylometry of Maoism: Quantifying the Language of Mao Zedong. In *Proceedings of the Joint 3rd International Conference on Natural Language Processing for Digital Humanities and 8th International Workshop on Computational Linguistics for Uralic Languages*, 76–81. Association for Computational Linguistics. https://aclanthology.org/2023.nlp4dh-1.9/.

Kurzynski, Maciej. 2024. Perplexity Games: Maoism vs. Literature through the Lens of Cognitive Stylometry. *Journal of Data Mining and Digital Humanities NLP4DH:23*. https://doi.org/10.46298/jdmdh.13131.

Kurzynski, Maciej. 2025. Poly-Temporal, Multi-Layered: A Techno-Cognitive Theory of Narrative Experience in Literature. *International Journal of Humanities and Arts Computing* 19 (1): 33–48. https://doi.org/10.3366/ijhac.2025.0343.

Kurzynski, Maciej, Xiaotong Xu, and Yu Feng. 2025. The Game of Keys and Queries: Parallelism and Cognitive Geometry in Chinese Regulated Verse. *International Journal of Humanities and Arts Computing* 19 (2): 143–157. https://doi.org/10.3366/ijhac.2025.0355.

Lee, Haiyan. 2014. *The Stranger and the Chinese Moral Imagination*. Stanford University Press.

Leese, Daniel. 2011. *Mao Cult: Rhetoric and Ritual in China's Cultural Revolution*. Cambridge University Press.

Leese, Daniel. 2014. A Single Spark: Origins and Spread of the Little Red Book in China. In *Mao's Little Red Book: A Global History*, edited by **Alexander C. Cook**, 23–42. Cambridge University Press.

Li, Tuo. 1998. 汪曾祺与现代汉语写作——兼谈毛文体 [Wang Zengqi and Modern Chinese Writing: Also on the Style of Mao's Writings]. 花城 [Huacheng].

Lifton, Robert J. 1956. Thought Reform of Chinese Intellectuals: A Psychiatric Evaluation. *The Journal of Asian Studies* 16, no. 1 (November): 75–88.

Lillicrap, Timothy P., Adam Santoro, Luke Marris, et al. 2020. Backpropagation and the Brain. *Nature Reviews Neuroscience* 21: 335–346. https://doi.org/10.1038/s41583-020-0277-3.

Link, Perry. 2012. Politics and the Chinese Language: What Mo-Yan's Defenders Get Wrong. Accessed: 2023-09-25. https://asiasociety.org/blog/asia/politics-and-chinese-language-what-mo-yans-defenders-get-wrong.

Link, Perry. 2013. *An Anatomy of Chinese: Rhythm, Metaphor, Politics*. Harvard University Press.

Liu, Jin, Edward Gunn, Fuad Youssef, et al. 2023. Fractality in Chinese Prose. *Digital Scholarship in the Humanities* 38, no. 2 (June): 604–620. https://doi.org/10.1093/llc/fqac062.

Lutosławski, Wincenty. 1898. Principes de Stylométrie Appliqués à la Chronologie des Oeuvres de Platon. *Revue des Études Grecques* 11 (41): 61–81.

Markov, Andrey. 2006. An Example of Statistical Investigation of the Text Eugene Onegin Concerning the Connection of Samples in Chains. *Science in Context* 19 (4): 591–600. https://doi.org/10.1017/S0269889706001074.

McDonald, Scott A., and Richard C. Shillock. 2004. Lexical Predictability Effects on Eye Fixations During Reading. In *The On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond*, edited by **Manuel Carreiras and Charles Clifton Jr.**, 77–94. Psychology Press.

Mitchell, Eric, Yoonho Lee, Alexander Khazatsky, et al. 2023. DetectGPT: Zero-Shot Machine-Generated Text Detection Using Probability Curvature. In *Proceedings of the 40th International Conference on Machine Learning*. ICML'23. Honolulu, Hawaii, USA. https://proceedings.mlr.press/v202/mitchell23a.html.

Mosteller, Frederick, and David L. Wallace. 1963. Inference in an Authorship Problem. *Journal of the American Statistical Association* 58 (302): 275–309. https://doi.org/10.2307/2283270.

O'Brien, Wendy. 2007. Telling Time: Literature, Temporality and Trauma. In *Temporality in Life as Seen Through Literature: Contributions to Phenomenology of Life*, edited by **Anna-Teresa Tymieniecka**, 86: 209–221. Analecta Husserliana. Springer. https://doi.org/10.1007/1-4020-5331-2_15.

OpenAI, Josh Achiam, Steven Adler, *et al.* 2023. GPT-4 Technical Report. arXiv preprint arXiv:2303.08774, https://arxiv.org/abs/2303.08774.

Paaß, Gerhard, and Sven Giesselbach. 2023. *Foundation Models for Natural Language Processing*. Springer. https://doi.org/10.1007/978-3-031-23190-2_2.

Pagnoni, Artidoro, Ramakanth Pasunuru, Pedro Rodriguez, et al. 2025. Byte Latent Transformer: Patches Scale Better Than Tokens. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, edited by **Wanxiang Che, Joyce Nabende, Ekaterina Shutova**, et al., 9238–9258. Association for Computational Linguistics, July. https://doi.org/10.18653/v1/2025.acl-long.453.

Pedinotti, Paolo, Eliana Di Palma, Ludovica Cerini, et al. 2021. A Howling Success or a Working Sea? Testing What BERT Knows about Metaphors. In *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP"*, edited by **Jasmijn Bastings, Yonatan Belinkov, Emmanuel Dupoux**, et al., 192–204. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.blackboxnlp-1.13.

Peer, Willie van, Paul Sopčák, Davide Castiglione, et al. 2021. Foregrounding. In *Handbook of Empirical Literary Studies*, edited by **Don Kuiken and Arthur M. Jacobs**, 145–176. De Gruyter. https://doi.org/10.1515/9783110645958-007.

Pimentel, Tiago, Ryan Cotterell, and Brian Roark. 2021. Disambiguatory Signals are Stronger in Word-initial Positions. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 31–41. Association for Computational Linguistics, April. https://doi.org/10.18653/v1/2021.eacl-main.3.

Qiu, Liben. 1999. ⟨⟨ 亚洲周刊 ⟩⟩ 评选百年中文小说百强 [**Asia Weekly's** Selection of the Top 100 Chinese Novels of the Century]. 当代文学研究资料与信息 [*Contemporary Literary Research Materials and Information*], no. 4.

Reagan, Andrew J, Lewis Mitchell, Dilan Kiley, et al. 2016. The Emotional Arcs of Stories are Dominated by Six Basic Shapes. *EPJ Data Science* 5: 31. https://doi.org/10.1140/epjds/s13688-016-0093-1.

Rebora, Simone. 2023. Sentiment Analysis in Literary Studies: A Critical Survey. *Digital Humanities Quarterly* 17 (2).

Resseguie, James L. 1991. Automatization and Defamiliarization in Luke 7: 36–50. *Literature and Theology* 5, no. 2 (June): 137–150.

Rybicki, Jan, Maciej Eder, and David L. Hoover. 2016. Computational Stylistics and Text Analysis. In *Doing Digital Humanities*, edited by **Constance Crompton, Richard J. Lane, and Ray Siemens**, 123–144. Routledge. https://doi.org/10.4324/9781315707860.

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press

Sayeau, Michael. 2018. Realism and the Novel. In *The Cambridge Companion to the Novel*, edited by **Eric Bulson**, 91–103. Cambridge University Press. https://doi.org/10.1017/9781316659694.007.

Schoenhals, Michael. 1992. *Doing Things with Words in Chinese Politics: Five Studies*. *Research Monograph* 41. Berkeley: Center for Chinese Studies, University of California at Berkeley.

Schoenhals, Michael. 2007. Demonising Discourse in Mao Zedong's China: People vs Non-People. *Totalitarian Movements and Political Religions* 8 (3): 465–482. https://doi.org/10.1080/14690760701571114.

Schuster, Sarah, Kim-Lara Weiss, Florian Hutzler, et al. 2025. Interactive and Additive Effects of Word Frequency and Predictability: A Fixation-related fMRI Study. *Brain and Language* 260:105508. https://doi.org/10.1016/j.bandl.2024.105508.

Shannon, Claude E. 1951. Prediction and Entropy of Printed English. *Bell System Technical Journal* 30 (1): 50–64. https://archive.org/details/bstj30-1-50.

Shen, Tianxiao, Tao Lei, Regina Barzilay, et al. 2017. Style Transfer from Non-Parallel Text by Cross-Alignment. *Advances in Neural Information Processing Systems (NeurIPS)*, 6830–6841. https://arxiv.org/abs/1705.09655.

Shklovsky, Viktor. 2017. *Viktor Shklovsky: A Reader*. Edited and translated by **Alexandra Berlina**. Bloomsbury Academic.

Sims, Matthew, Jong Ho Park, and David Bamman. 2019. Literary Event Detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3623–3634. Association for Computational Linguistics. https://aclanthology.org/P19-1353/.

Snow, Don. 2004. *Cantonese as Written Language: The Growth of a Written Chinese Vernacular*. Hong Kong University Press.

Soni, Ansh, Sudhanshu Srivastava, Konrad Kording, et al. 2024. Conclusions About Neural Network to Brain Alignment Are Profoundly Impacted by the Similarity Measure. bioRxiv, https://doi.org/10.1101/2024.08.07.607035.

Steiner, Peter. 1984. *Russian Formalism: A Metapoetics*. Cornell University Press.

Sun, Simeng, Katherine Thai, and Mohit Iyyer. 2022. CHAPTERBREAK: A Challenge Dataset for Long-Range Language Models. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3704–3714. Association for Computational Linguistics, July. https://doi.org/10.18653/v1/2022.naacl-main.271.

Teodorescu, Daniela, and Saif Mohammad. 2023. Evaluating Emotion Arcs Across Languages: Bridging the Global Divide in Sentiment Analysis. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, edited by **Houda Bouamor, Juan Pino, and Kalika Bali**, 4124–4137. Association for Computational Linguistics, December. https://doi.org/10.18653/v1/2023.findings-emnlp.271.

Tikochinski, Refael, Ariel Goldstein, Yoav Meiri, et al. 2025. Incremental accumulation of linguistic context in artificial and biological neural networks. *Nature Human Behaviour* 16 (803).

Tsur, Reuven. 2008. *Toward a Theory of Cognitive Poetics: Second , Expanded & Updated Edition*. Sussex Academic Press.

Varnum, Michael E. W., Nicolas Baumard, Mohammad Atari, et al. 2024. Large language models based on historical text could offer informative tools for behavioral science. *Proceedings of the National Academy of Sciences* 121 (42): e2407639121. https://doi.org/10.1073/pnas.2407639121.

Vos, Ruby de. 2024. Making the Strange Familiar: Getting Intimate with Toxicity. In *E(n)stranged: Rethinking Defamiliarization in Literature and Visual Culture*, edited by **Nilgun Bayraktar and Alberto Godioli**, 73–92. Palgrave Macmillan, October. https://doi.org/10.1007/978-3-031-60859-9_4.

Wake, Paul. 2013. Narrative and Narratology. In *The Routledge Companion to Critical and Cultural Theory, 2nd*, edited by **Simon Malpas and Paul Wake**, 34–45. Routledge.

Weidinger, Laura, John F. J. Mellor, Maribeth Rauh, et al. 2021. Ethical and Social Risks of Harm from Language Models. ArXiv, https://arxiv.org/abs/2112.04359.

Wu, Kangxi, Liang Pang, Huawei Shen, et al. 2023. LLMDet: A Third Party Large Language Models Generated Text Detection Tool. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 2113–2133. Singapore. Association for Computational Linguistics. https://aclanthology.org/2023.findings-emnlp.139/.

Wu, Yaru, Yuri Bizzoni, Pascale Moreira, et al. 2024. Perplexing Canon: A Study on GPT-Based Perplexity of Canonical and Non-Canonical Literary Works. In *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)*, 172–184. Association for Computational Linguistics, March. https://aclanthology.org/2024.latechclfl-1.16/.

Xiang, Rong, Emmanuele Chersoni, Yixia Li, et al. 2025. Cantonese Natural Language Processing in the Transformers Era: A Survey and Current Challenges. *Language Resources and Evaluation* 59, 1747–1773. https://doi.org/10.1007/s10579-024-09744-w.

Yan, Lixiang, Lele Sha, Linxuan Zhao, et al. 2024. Practical and Ethical Challenges of Large Language Models in Education: A Systematic Scoping Review. *British Journal of Educational Technology* 55 (1): 90–112. https://doi.org/10.1111/bjet.13370.

Yeung, Wayne C. F. 2023. The Concept of the Cantophone: Memorandum for a Stateless Literary History. *Sino-Platonic Papers* 334 (June): 1–49.

Yu, Yijiong, Ziyun Dai, Zekun Wang, et al. 2025. OpenCSG Chinese Corpus: A Series of High-quality Chinese Datasets for LLM Training. *arXiv preprint arXiv:2501.08197*, https://arxiv.org/abs/2501.08197.

Zhao, Xuandong, Prabhanjan Ananth, Lei Li, et al. 2023. Provable Robust Watermarking for AI-Generated Text. arXiv preprint arXiv:2306.17439, https://arxiv.org/abs/2306.17439.

https://doi.org/10.1017/chr.2025.10020 Published online by Cambridge University Press