

Phase transition on a context-sensitive random language model with short-range interactions

Yuma Toji,¹ Jun Takahashi², Vwani Roychowdhury³, and Hideyuki Miyahara^{1,4,*}¹*Division of Information Science and Technology, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido 060-0814, Japan*²*Institute for Solid State Physics, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8581, Japan*³*Department of Electrical and Computer Engineering, Henry Samueli School of Engineering and Applied Science, University of California, Los Angeles, California 90095, USA*⁴*School of Computing, Institute of Science Tokyo, Ookayama, Tokyo 152-8550, Japan*

(Received 31 March 2026; accepted 21 July 2026; published 19 October 2026)

Since the random language model was proposed by De Giuli [*Phys. Rev. Lett.* **122**, 128301 (2019)], language models have been investigated intensively from the viewpoint of statistical mechanics. Recently, the existence of a Berezinskii-Kosterlitz-Thouless transition was numerically reported in models with long-range interactions between symbols. In statistical mechanics, it has long been known that long-range interactions can induce phase transitions. Therefore, it has remained unclear whether phase transitions observed in language models originate from genuinely linguistic properties that are absent in conventional spin models. In this study, we construct a random language model with short-range interactions and numerically investigate its statistical properties. Our model belongs to the class of context-sensitive grammars in the Chomsky hierarchy and allows explicit reference to contexts. We find that a phase transition occurs even when the model refers only to contexts whose length remains constant with respect to the sentence length. This result indicates that finite-temperature phase transitions in language models are genuinely induced by the intrinsic nature of language, rather than by long-range interactions.

DOI: [10.1103/vnf1-mfrs](https://doi.org/10.1103/vnf1-mfrs)

I. INTRODUCTION

Phase transitions, such as the liquid-gas transition, metal-insulator transition, magnetic phase transition, and superconducting transition, are ubiquitous phenomena in statistical mechanics, arising from collective behavior in many-body systems [1–3]. Most systems around us can be regarded as many-body systems because they involve interactions among their constituent elements. It is therefore natural to expect that phase-transition-like phenomena may also appear beyond traditional physical systems. In recent years, phenomena analogous to phase transitions have been reported in natural language processing and have attracted considerable attention, suggesting a deep connection between linguistic structure and statistical physics. In particular, emergent abilities and scaling behaviors in large language models [4,5] have been interpreted as manifestations of such collective effects. Notably, prior to these striking findings, De Giuli formulated the random language model, which describes language generation as a statistical ensemble over grammars, and argued for the existence of phase transitions in language models [6], leading to further discussions [7,8]. In this framework, sentences

are generated probabilistically according to randomly drawn grammar rules, enabling linguistic structure to be analyzed using tools from statistical mechanics. This model establishes a concrete bridge between formal language theory and many-body physics, opening the possibility of studying linguistic phenomena in terms of thermodynamic phases and critical behavior.

Building upon this framework, we previously investigated a variant of the random language model incorporating long-range interactions between symbols [9]. Through numerical simulations, we provided evidence suggesting the presence of a Berezinskii-Kosterlitz-Thouless (BKT) transition [10–14] in this system, revealing nontrivial critical behavior analogous to that observed in low-dimensional statistical-mechanical models. This result provided strong evidence that language models can exhibit genuine phase transitions. However, long-range interactions are well known to induce phase transitions even in otherwise simple spin systems [13,15–19]. Therefore, it remains unclear whether the observed phase transitions in language models genuinely reflect intrinsic properties of language, or whether they are merely consequences of long-range couplings.

To address this fundamental question, it is natural to construct a theoretical model that eliminates long-range interactions and to isolate the mechanism of interest behind linguistic statistical regularities. To this end, we construct a random language model that incorporates only short-range interactions. Specifically, we introduce a generative model in analogy with the one-dimensional Potts model, where interactions are restricted to finite-range contexts whose

*Contact author: miyahara@ist.hokudai.ac.jp;
hmiyahara512@gmail.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

length remains constant with respect to the sentence length [20]. Sentences are generated from this model and analyzed using standard statistical-mechanical observables, including magnetization, susceptibility, Binder cumulant, and correlation functions. Our numerical analysis reveals that a phase transition occurs even in the absence of long-range interactions. Remarkably, the transition appears to be consistent with the BKT-like transition observed in the previously studied long-range model. In this transition, correlation functions exhibit power-law decay not only at the critical point but throughout an extended critical phase below it, in contrast to conventional first- or second-order phase transitions. These findings demonstrate that nontrivial critical phenomena in language models do not rely on long-range couplings, but instead emerge from structural properties intrinsic to the linguistic generative process.

This paper is organized as follows. In Sec. II, we explain the details of the proposed model. The physical quantities calculated from the model are described in Sec. III. The results of the representative parameters are shown in Sec. IV, and discussed in Sec. V. Finally, we conclude this paper in Sec. VI.

II. PROPOSED MODEL

In this paper, we consider the model that has the following three generative rules [9,21–24]:

$$X \rightarrow x \quad (\text{probability: } qt), \quad (2.1a)$$

$$X \rightarrow YZ \quad (\text{probability: } q(1-t)), \quad (2.1b)$$

$$Z_- X Z_+ \rightarrow Z_- Y Z_+ \quad (\text{probability: } (1-q)), \quad (2.1c)$$

where the lowercase and uppercase letters represent terminal and nonterminal symbols, respectively.

We elaborate on the definitions of three rules in Eq. (2.1), respectively. Rule (2.1a) transforms a nonterminal symbol into its *corresponding* terminal symbol. Here, *corresponding* means that there is a one-to-one relation between a nonterminal symbol A_k and a terminal symbol a_k , so a production rule like $A_2 \rightarrow a_4$ does not exist. Thus, the numbers of terminal and nonterminal symbols are specified by K ; that is, the model handles K types of terminal symbols $a_1, a_2, \dots, a_K$ and K types of nonterminal symbols $A_1, A_2, \dots, A_K$.

Rule (2.1b) transforms a nonterminal symbol into two nonterminal symbols; in other words, this rule increases the number of nonterminal symbols in the sentence. This rule has a parameter $\epsilon \in [0, 1]$, which controls the tendency of what symbols appear on the right-hand side of rule (2.1b). Specifically, when the symbol on the left-hand side of rule (2.1b) is A_k , the symbols Y and Z on the right-hand side of rule (2.1b) are selected as follows:

$$Y, Z = \begin{cases} A_k, & (\text{probability: } 1 - \frac{(K-1)\epsilon}{K}), \\ A \in \Sigma_{\text{nt}} \setminus \{A_k\}, & (\text{probability: } \frac{\epsilon}{K}). \end{cases} \quad (2.2)$$

Note that $\Sigma_{\text{nt}} = \{A_1, A_2, \dots, A_K\}$ denotes the set of nonterminal symbols. According to this formulation, Y and Z are always selected as X when $\epsilon = 0$, and selected uniformly when $\epsilon = 1$. The model that consists of only rules (2.1a) and (2.1b) corresponds to a context-free grammar (CFG).

As discussed below, our model is identical to the one-dimensional Potts model with short-range interactions if rule

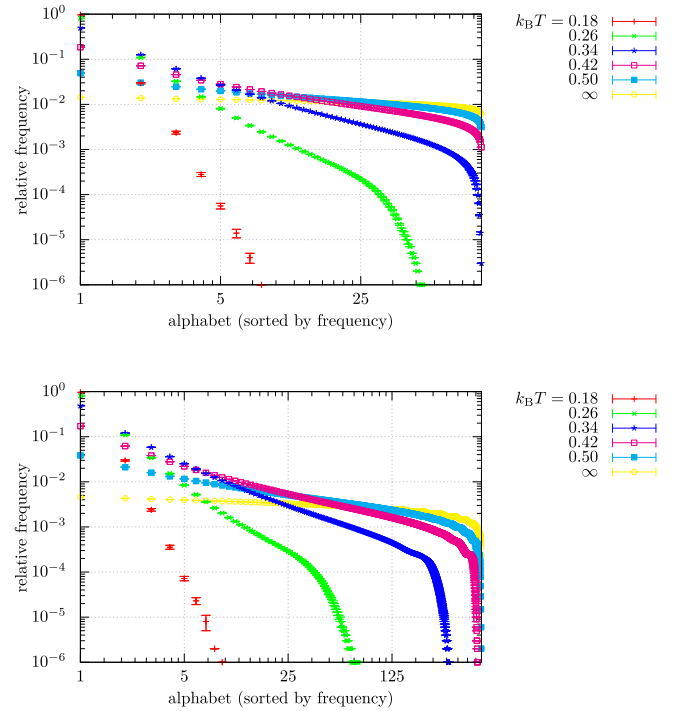


FIG. 1. Ranking of symbol appearance frequencies for (upper) $K = 100$ and (lower) $K = 500$, plotted on log-log scales. The critical temperature in both cases is $k_B T \simeq 0.42$. We set $J = 1.0$, $q = 10^{-1.0}$, $t = 0.0$, and $\epsilon = 0.00$. The curves for different $k_B T$ are overlaid. Each curve represents the statistical average over a sample size of 6400 realizations, with error bars denoting the standard errors.

(2.1b) is removed. Therefore, rule (2.1b), which represents a fundamental property of formal language, is not considered in conventional physical models. Since the Potts model, which does not have rule (2.1b), does not exhibit a finite-temperature phase transition, the symbol branching (generation) process based on a formal language model might be capable of explaining complex phenomena (the existence of a phase transition and polynomial decay of correlation functions) that are ubiquitously observed in natural languages.

Rule (2.1c) characterizes the context sensitivity of the model. This rule replaces the symbol X with the symbol Y , depending on the Z_- and Z_+ , the symbols before and after X . If X is A_k , rule (2.1c) first selects a destination symbol Y from the set of nonterminal symbols other than A_k . Suppose Y is A_j . Then, the probability p of applying rule (2.1c) is calculated as follows:

$$p = \min \left(1, \exp \left(-\frac{\Delta E}{k_B T} \right) \right), \quad (2.3a)$$

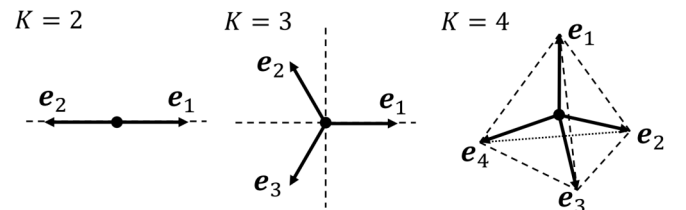


FIG. 2. A schematic diagram of the vectors e_k for $K = 2, 3$, and 4 . They sum to the zero vector, and the inner products of any two different vectors are constant.

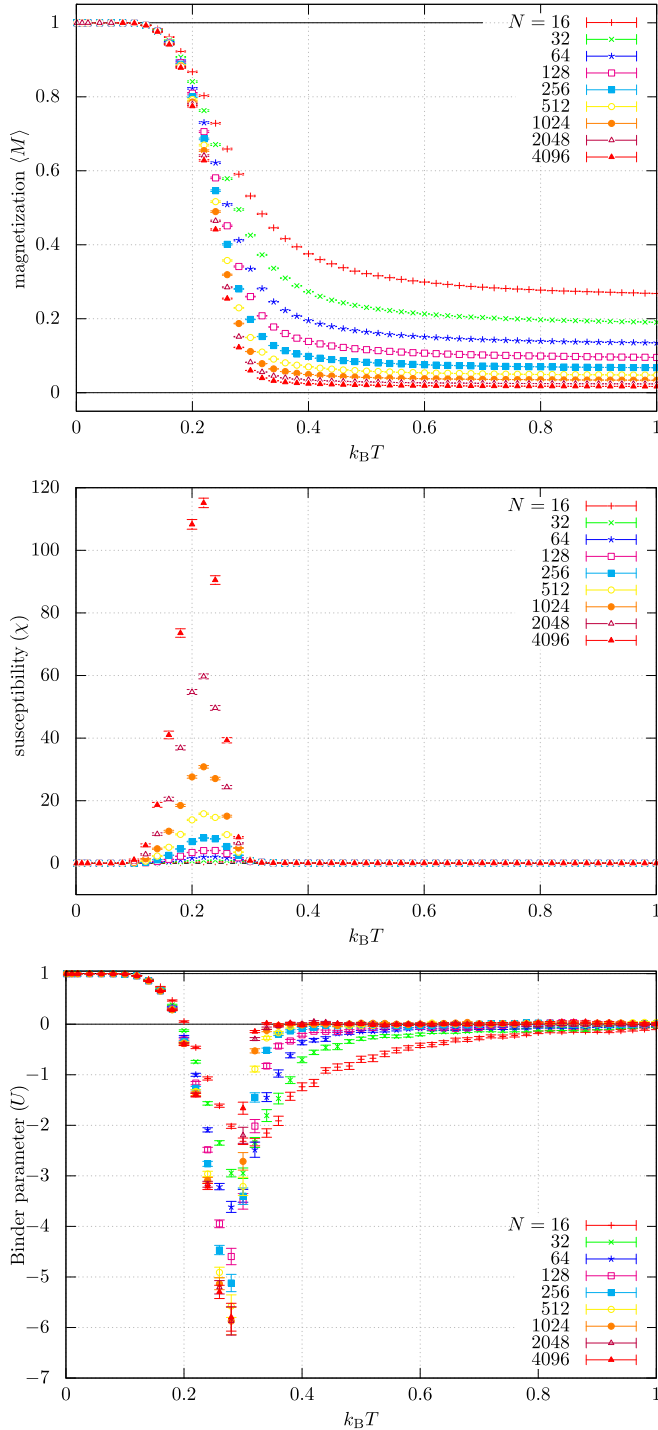


FIG. 3. Temperature dependence of (upper) the magnetization [Eq. (3.1)], (middle) the susceptibility [Eq. (3.4)] and (lower) the Binder parameter [Eq. (3.5)]. We set $K = 20, J = 1.0, q = 10^{-2.0}, t = 0.0$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

$$\Delta E := J(\delta_{\sigma_i, \sigma_{i-1}} - \delta_{\tilde{\sigma}_i, \tilde{\sigma}_{i-1}} + \delta_{\sigma_i, \sigma_{i+1}} - \delta_{\tilde{\sigma}_i, \tilde{\sigma}_{i+1}}). \quad (2.3b)$$

Here, $k_B T$ is the temperature parameter of rule (2.1c) and the parameter J is a positive constant. The variable $\sigma_i \in \{1, 2, \dots, K\}$ denotes the i th symbol in the sentence. If the

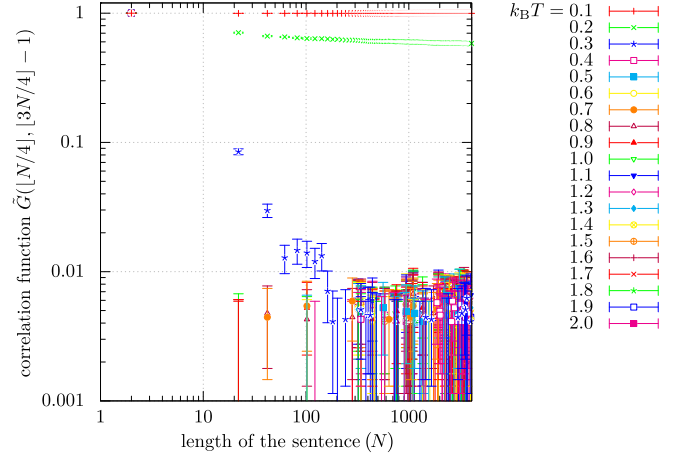


FIG. 4. The correlation function, [Eq. (3.7)] with $i = \lfloor N/4 \rfloor$ and $j = \lfloor 3N/4 \rfloor - 1$. We set $K = 20, J = 1.0, q = 10^{-2.0}, t = 0.0$, and $\epsilon = 0.00$. We show the results for various $k_B T = 0.1, 0.2, \dots, 2.0$, and the curves for different $k_B T$ are overlaid.

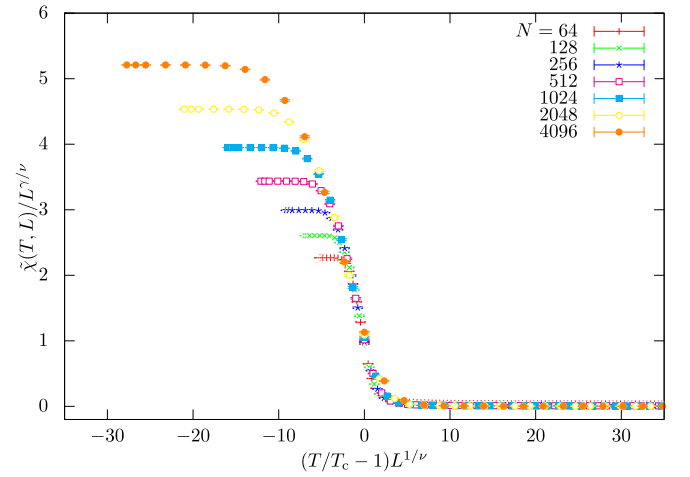


FIG. 5. Finite-size scaling of $\tilde{\chi}$ with $T_c = 0.24, \nu = 2.50$, and $\gamma = 2.00$. We set $K = 20, J = 1.0, q = 10^{-2.0}, t = 0.0$, and $\epsilon = 0.00$. We varied $N = 64, 128, \dots, 4096$.

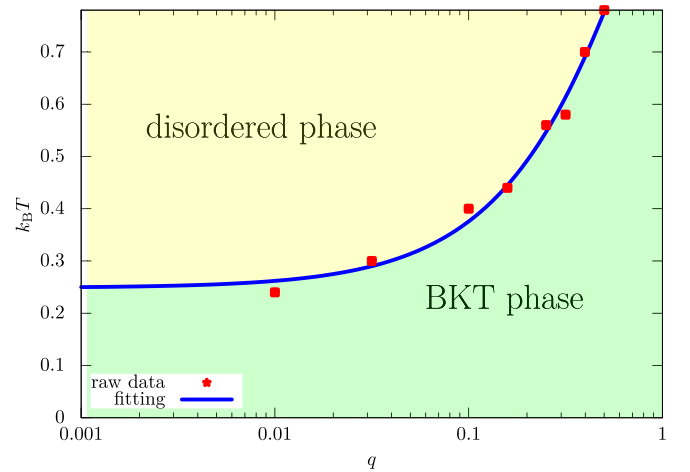


FIG. 6. Phase diagram for $K = 20, J = 1.0, t = 0.0$, and $\epsilon = 0.00$. The horizontal and vertical axes represent the parameter q , which controls the probability of selecting production rules, and the critical temperature $k_B T$, respectively.

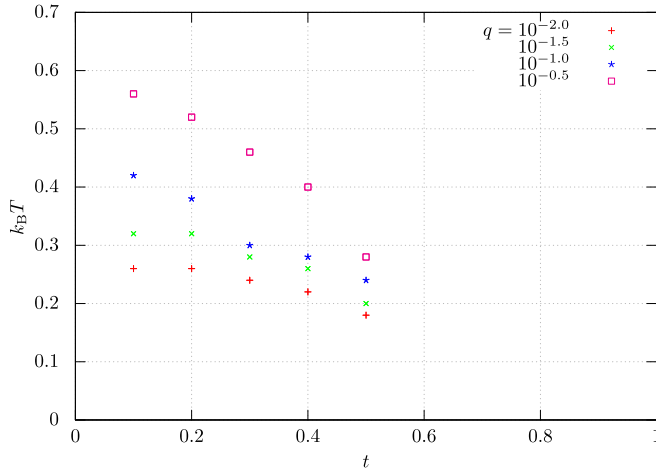


FIG. 7. Phase diagram for $K = 20$, $J = 1.0$, and $\epsilon = 0.00$. The horizontal and vertical axes represent the parameter t , which controls the probability of selecting production rules, and the critical temperature $k_B T$, respectively.

i th symbol is a_k or A_k , then σ_i is equal to k . Regarding the physical meaning of the temperature parameter T , when T is small, replacements based on the contexts tend to occur, whereas when T is large, random replacements tend to occur.

This process is the same as the one used in the simulation of the K -state Potts model by the Metropolis-Hastings method with single-spin flips [25–27]. In other words, this model adopts the interactions of the Potts model as interactions between symbols in the sentences. However, it should be noted that Eq. (2.3a) is not introduced as a Markov chain Monte Carlo sampling algorithm for generating equilibrium configurations. Since the state space expands as the sentence length grows, the system is not ergodic. Rather, Eq. (2.3a) introduces the local contextual dependence of natural language as a simple rule.

Our model is a highly simplified model of language. Nevertheless, when we increase K , the number of possible states for each symbol, we confirm that the model reproduces phe-

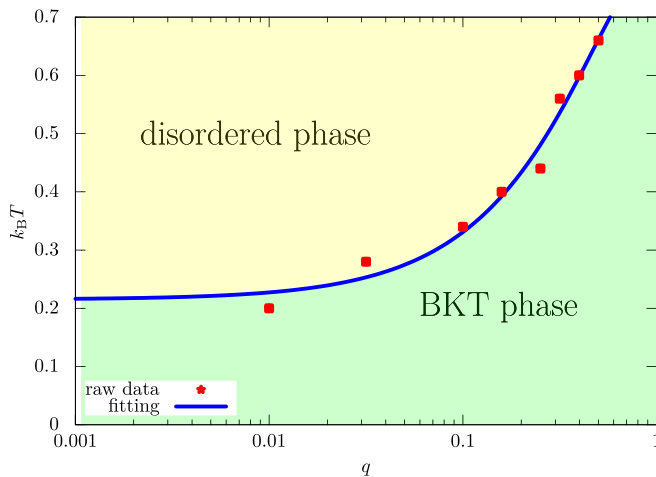


FIG. 8. Phase diagram for $K = 2$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. The horizontal and vertical axes represent the parameter q , which controls the probability of selecting production rules, and the critical temperature $k_B T$, respectively.

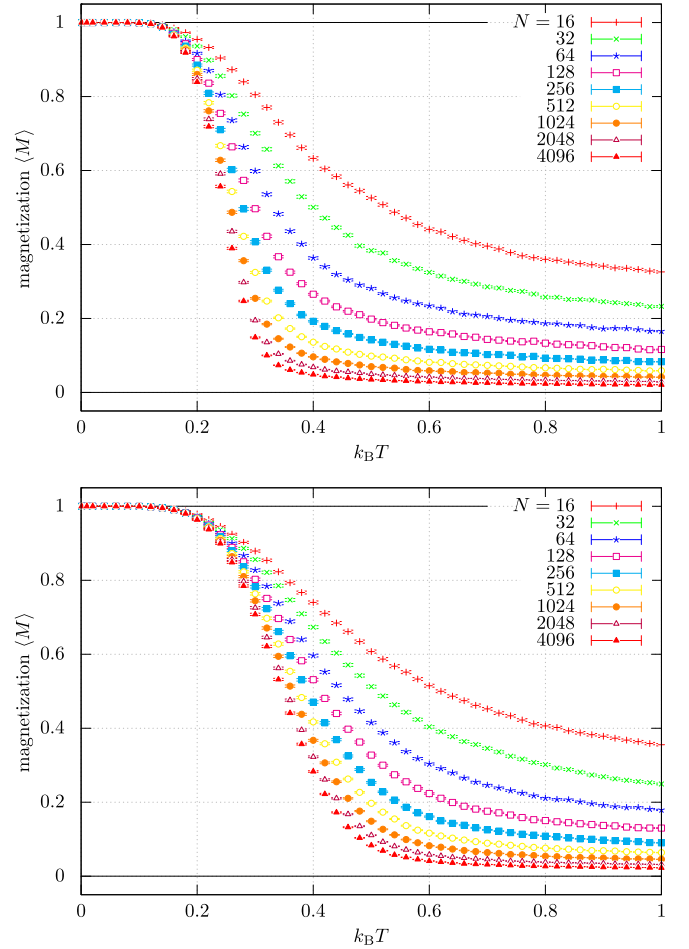


FIG. 9. Temperature dependence of the magnetization [Eq. (3.1)]. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 2$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

nomena similar to Zipf’s law observed in natural languages. Figure 1 shows the relative frequencies of symbols, ranked in descending order and plotted on log-log axes, for $K = 100$ and $K = 500$. Although the model is based on the strong assumption that all symbols are treated completely symmetrically, the fact that it can reproduce Zipf-like behavior provides supporting evidence for the validity of the results obtained from this model. Considering that power-law behavior is a typical example of critical phenomena in phase transitions, one might naturally connect it to the observed Zipf-like behavior; however, their precise relationship remains unclear.

TABLE I. The critical temperature of each parameter set. The symbol “–” means no phase transitions occur in the parameter setting.

t	0.3	0.5	0.7
$q = 10^{-2.0}$	0.30	0.26	–
$q = 10^{-1.0}$	0.36	0.28	–

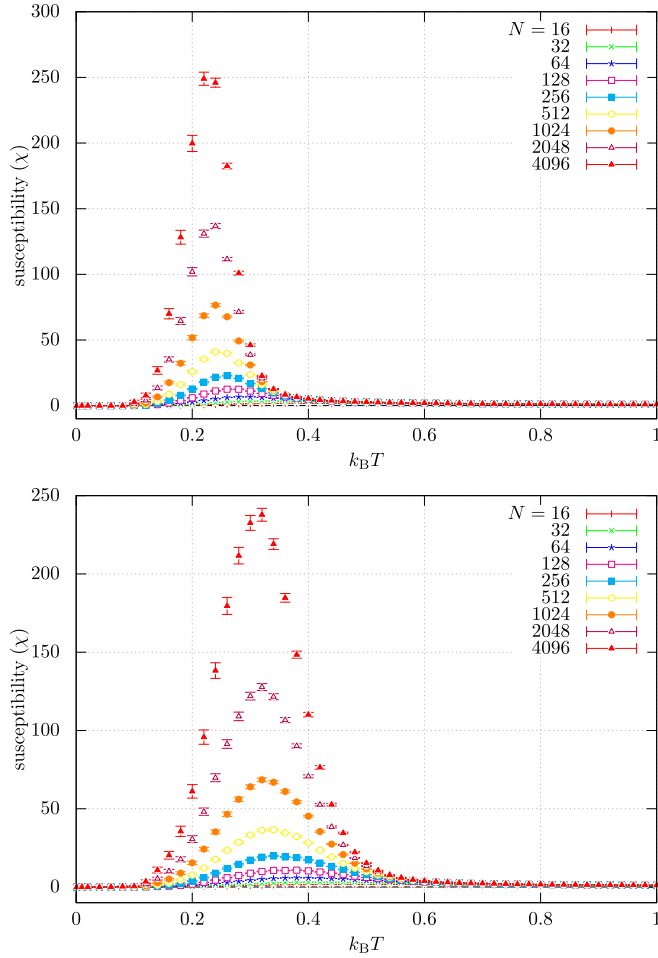


FIG. 10. Temperature dependence of the susceptibility [Eq. (3.4)]. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 2, J = 1.0, t = 0.0$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

From the viewpoint of formal language theory, the present model belongs to the class of context-sensitive grammars in the Chomsky hierarchy, whereas the random language model introduced by De Giuli corresponds to a context-free grammar [28]. In this sense, our construction can be regarded as a natural extension of De Giuli's framework, incorporating explicit contextual dependence while remaining within a probabilistic generative setting.

III. PHYSICAL QUANTITIES

In this section, we describe the physical quantities calculated from the model explained in Sec. II. As an order parameter, we used the magnetization M shown in Eq. (3.1) that is defined in the same way as in the Potts model [20]:

$$M := \| \mathbf{M} \|, \quad (3.1)$$

where

$$\mathbf{M} := \frac{1}{N} \sum_{i=1}^N \mathbf{e}_{\sigma_i}. \quad (3.2)$$

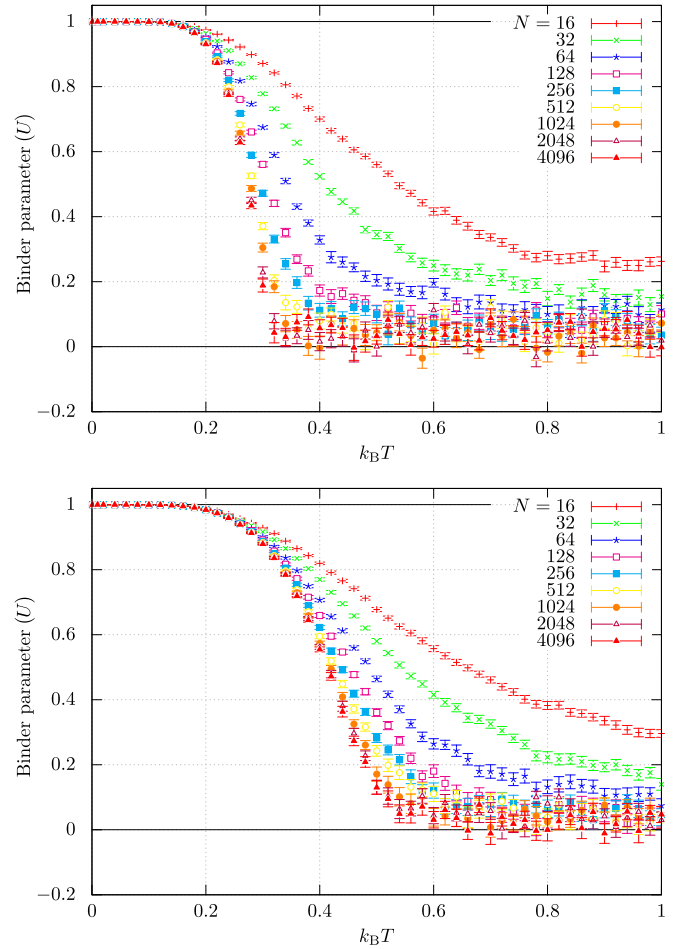


FIG. 11. Temperature dependence of the Binder parameter [Eq. (3.5)]. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 2, J = 1.0, t = 0.0$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

Here, N denotes the length of the sentence. The vectors $\{\mathbf{e}_k\}_{k=1}^K$ satisfy the following conditions:

$$\sum_{k=1}^K \mathbf{e}_k = \mathbf{0}, \quad (3.3a)$$

$$\mathbf{e}_k \cdot \mathbf{e}_l = \begin{cases} 1, & (k = l), \\ \text{const}, & (k \neq l). \end{cases} \quad (3.3b)$$

Equation (3.3a) leads to $M = 0$ in Eq. (3.1) for a paramagnetic state and Eq. (3.3b) is a condition that reflects the system's symmetry in physical quantities. The vectors $\mathbf{e}_k$ for $K = 2, 3$, and 4 are illustrated in Fig. 2.

The magnetic susceptibility χ is defined as the variance of the magnetization M :

$$\chi := N(\langle M^2 \rangle - \langle M \rangle^2). \quad (3.4)$$

The susceptibility, Eq. (3.4), is divergent at a critical point; thus, it is quite important to see its behavior to discuss a phase transition via numerical simulations.

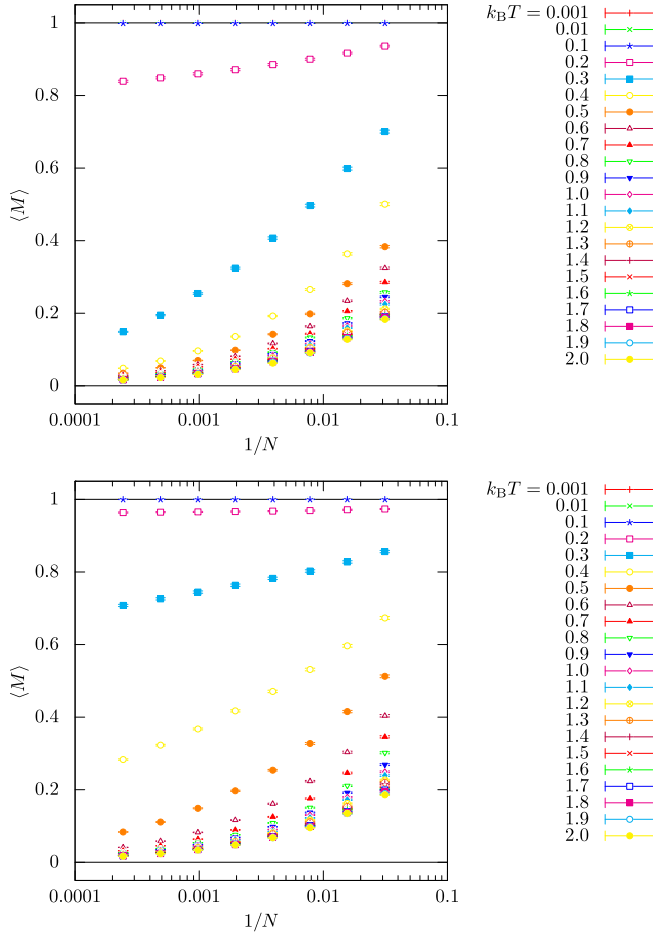


FIG. 12. The system-size dependence of the magnetization [Eq. (3.1)]. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 2$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various $k_B T = 0.001, 0.01, 0.1, 0.2, \dots, 2.0$, and the curves for different $k_B T$ are overlaid.

The Binder parameter U is the kurtosis of the distribution of M , normalized such that U equals 0 when the distribution of the magnetization is Gaussian and equals 1 when the distribution is a delta function:

$$U := -\frac{K-1}{2} \left(\frac{\langle M^4 \rangle}{\langle M^2 \rangle^2} - \frac{K+1}{K-1} \right). \quad (3.5)$$

The Binder parameter plays an important role in distinguishing between the standard second-order phase transition and the BKT transition [29,30].

The correlation function is also an important quantity to discuss phase transition since it typically shows a power-law decay at a critical point. We define it as follows [31]:

$$G_{k,l} := \langle \mathbf{e}_{\sigma_k} \cdot \mathbf{e}_{\sigma_l} \rangle - \langle \mathbf{e}_{\sigma_k} \rangle \cdot \langle \mathbf{e}_{\sigma_l} \rangle. \quad (3.6)$$

To observe the power-law decay of the correlation function associated with critical phenomena, it is sufficient to evaluate the following quantity:

$$\tilde{G}_{k,l} := \langle \mathbf{e}_{\sigma_k} \cdot \mathbf{e}_{\sigma_l} \rangle. \quad (3.7)$$

In this paper, we compute Eq. (3.7) instead of Eq. (3.6).

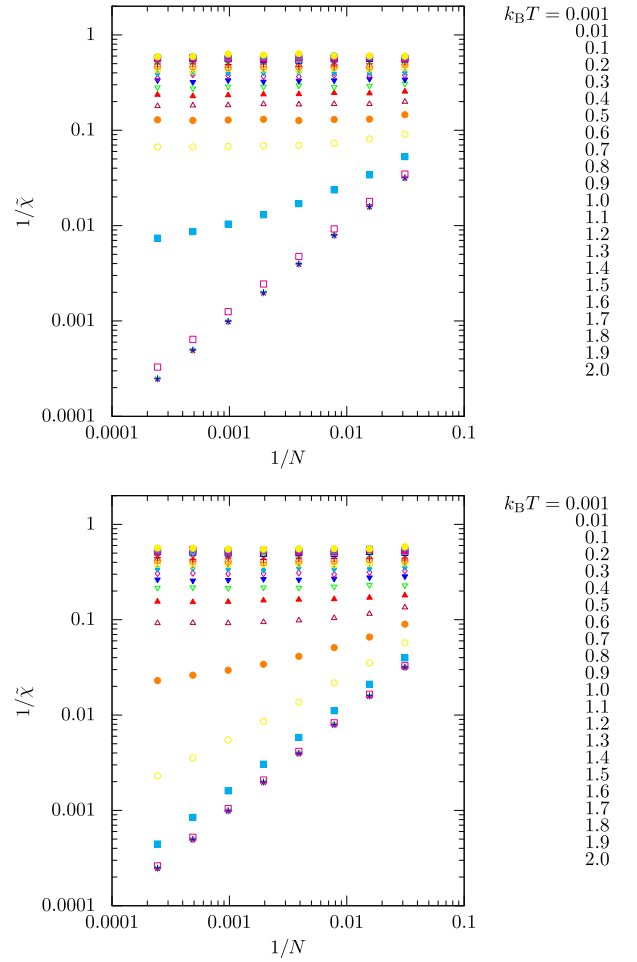


FIG. 13. The system-size dependence of the susceptibility [Eq. (3.4)]. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 2$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various $k_B T = 0.001, 0.01, 0.1, 0.2, \dots, 2.0$, and the curves for different $k_B T$ are overlaid.

Finally, we describe the finite-size scaling [32–35]

$$\frac{\tilde{\chi}(T, N)}{N^{\frac{\gamma}{\nu}}} = \tilde{f}_{\tilde{\chi}}(N^{\frac{1}{\nu}} t). \quad (3.8)$$

Here, $t := (T - T_c)/T_c$ and $\tilde{\chi}$ is defined as

$$\tilde{\chi} := N \langle M^2 \rangle. \quad (3.9)$$

The finite-size scaling relation holds around the critical temperature. When we plot $\tilde{\chi}(T, N)/N^{\gamma/\nu}$ against $N^{1/\nu} t$ for different system sizes N , the data collapse onto a single universal curve $\tilde{f}$.

IV. NUMERICAL SIMULATIONS

In this section, we present the results for $K = 20$, $\epsilon = 0.00$ as representative cases. All physical quantities are calculated as statistical averages over 6400 realizations, and the error bars shown in the figures represent the standard errors. While the magnetization M and the correlation function $\tilde{G}_{k,l}$ can be averaged directly across all 6400 realizations, the magnetic susceptibility χ and the Binder parameter U require ensemble

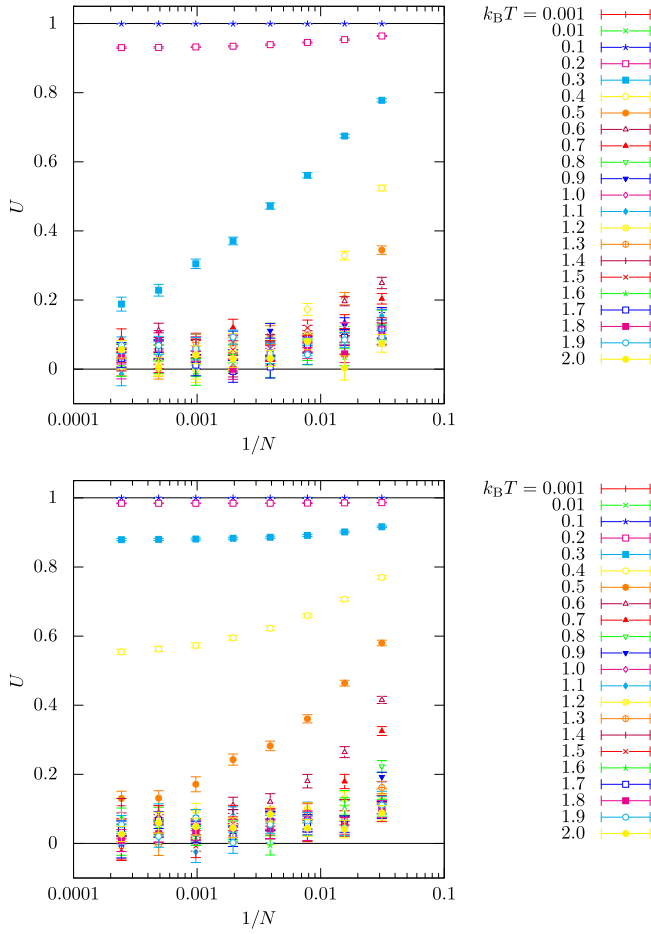


FIG. 14. The system-size dependence of the Binder parameter [Eq. (3.5)]. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 2$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various $k_B T = 0.001, 0.01, 0.1, 0.2, \dots, 2.0$, and the curves for different $k_B T$ are overlaid.

averages of magnetization within their definitions. Therefore, we divide the 6400 realizations into 80 independent subsets, with each subset consisting of 80 realizations. A single value of χ or U is calculated for each subset, providing 80 data points for these two quantities are computed from these 80 values.

Figure 3 shows the magnetization, susceptibility, and Binder parameter for $K = 20$, $t = 0.0$, $q = 10^{-2.0}$, and $\epsilon = 0.00$. For $t = 0.0$, we can consider a simple thermodynamic limit, since the production rule (2.1a) is not chosen, and the sentence continues to extend indefinitely.

The magnetization and susceptibility show singularity and divergence, respectively, around $k_B T \sim 0.24$. The Binder parameter also changes from 0 to negative values around $k_B T \sim 0.24$, and then approaches 1 as the temperature decreases.

Figure 4 shows the correlation function. The plots for $k_B T = 0.1$ and 0.2 (i.e., $k_B T < 0.24$) exhibit a nearly linear behavior, whereas those for $k_B T > 0.24$ decay rapidly.

Figure 5 shows the result of the finite-size scaling, in which the critical exponents are determined via the grid-search method.

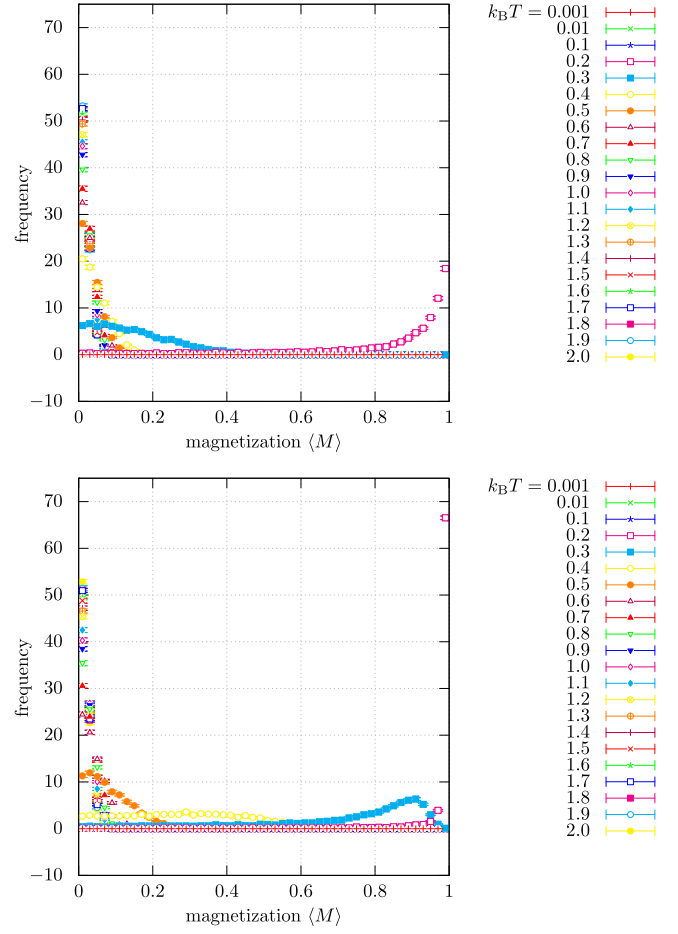


FIG. 15. The histogram of the magnetization. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 2$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various $k_B T = 0.001, 0.01, 0.1, 0.2, \dots, 2.0$, and the curves for different $k_B T$ are overlaid.

In Figs. 3–5, we fixed $q = 10^{-2.0}$ and $t = 0.0$; to see the q -, t -dependence of the model, Eq. (2.1), we show phase diagrams in Figs. 6 and 7 estimating the critical temperatures for various values of q and t . According to the forms of the Binder parameter in Fig. 3 (lower) and the correlation function in Fig. 4, these numerical results strongly suggest that BKT-like phase is realized, as shown in Fig. 6 [9]. Furthermore, we see an interesting phenomenon in Fig. 7, that is, the sudden drops of the critical temperature at $t = 0.5$. These points will be elaborated more in the following section.

V. DISCUSSIONS

All three panels in Fig. 3 show the existence of a phase transition. Although the singularity of the magnetization and the divergence of the susceptibility are weak due to finite-size effects, we can see that these tendencies become stronger as the system size N increases. The correlation function is shown in Fig. 4. In conventional second-order phase transitions, it is known that the correlation function decays polynomially with distance at the transition temperature. However, in this figure, the correlation function decays polynomially below the transition temperature $k_B T_* = 0.24$. Thus, critical phenomena

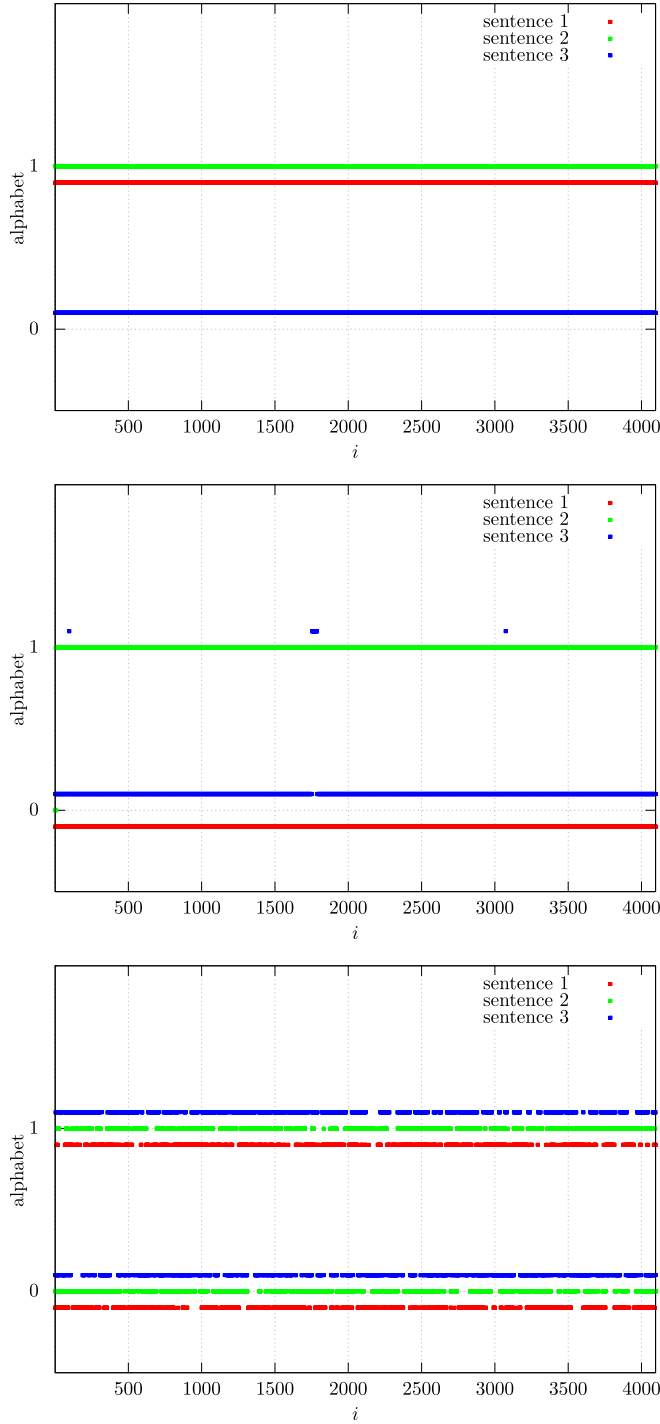


FIG. 16. Typical configurations of symbols. The upper panel shows an example sentence at a temperature below the transition ($k_B T = 0.001$) with magnetization $0.98 \leq M < 1.00$. The middle panel shows an example sentence at a temperature just above the transition ($k_B T = 0.20$) with magnetization $0.98 \leq M < 1.00$. The lower panel shows an example sentence at a temperature above the transition ($k_B T = 0.40$) with magnetization $0.00 \leq M < 0.02$. We set $K = 2$, $J = 1.0$, $q = 10^{-2.0}$, $t = 0.0$, and $\epsilon = 0.00$.

occur not only at the critical temperature but also below it. This characteristic is highly consistent with the properties of the BKT transition.

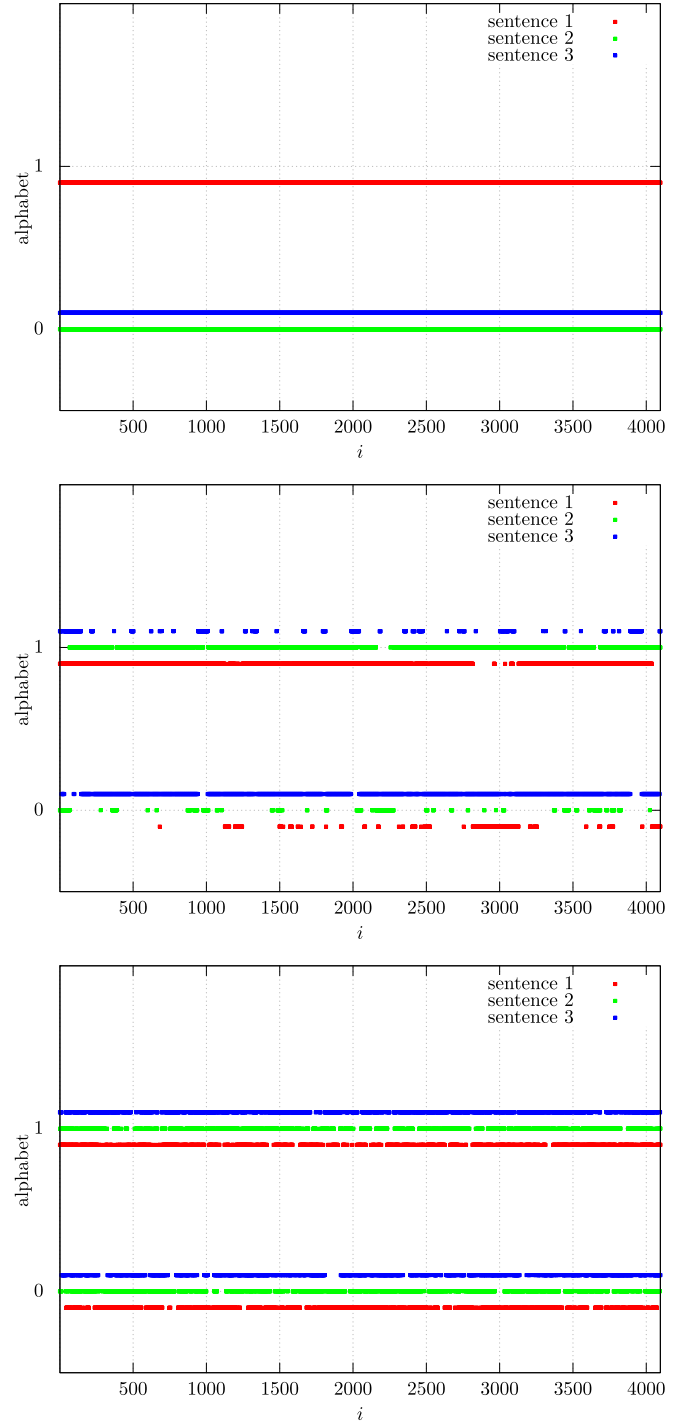


FIG. 17. Typical configurations of symbols. The upper panel shows an example sentence at a temperature below the transition ($k_B T = 0.14$) with magnetization $0.98 \leq M < 1.00$. The middle panel shows an example sentence at a temperature just above the transition ($k_B T = 0.34$) with magnetization $0.72 \leq M < 0.74$. The lower panel shows an example sentence at a temperature above the transition ($k_B T = 0.54$) with magnetization $0.00 \leq M < 0.02$. We set $K = 2$, $J = 1.0$, $q = 10^{-1.0}$, $t = 0.0$, and $\epsilon = 0.00$.

If the observed transition indeed belongs to the BKT universality class, an important question is how such behavior can emerge in a system with only short-range interactions.

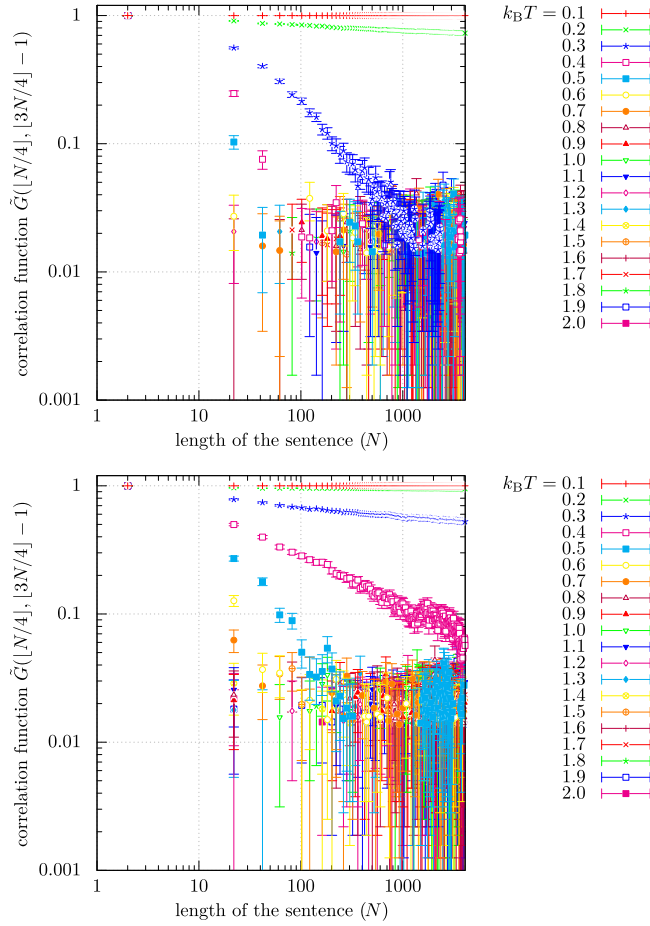


FIG. 18. The correlation function, [Eq. (3.7)] with $i = \lfloor N/4 \rfloor$ and $j = \lfloor 3N/4 \rfloor - 1$. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 2, J = 1.0, t = 0.0$, and $\epsilon = 0.00$. We show the results for various $k_B T = 0.1, 0.2, \dots, 2.0$, and the curves for different $k_B T$ are overlaid.

The BKT transition is known as a phenomenon that occurs in the XY model, but it is also known to occur in one-dimensional models with long-range interactions [13]. However, our model has only short-range interactions. We consider that the key to resolving this controversy is the process of increasing spin variables, which is unique to language models. We hypothesize that the key to resolving this controversy is the process of increasing spin variables, which is unique to language models. Even if the two spins do not interact when they are far apart in the large system, they may have interacted when the system size was still small, that is, at the beginning of the sentence generation, and this history may remain. While this interpretation remains hypothetical at present, such a mechanism could qualitatively explain the emergence of effective long-range interactions from purely short-range rules. Further numerical or analytical verification of this generation-history effect is left for future work.

Finally, Fig. 7 shows the $k_B T-t$ phase diagram. We show it for various values of q . For all q , phase transitions occur at finite temperature below $t = 0.5$, but no phase transitions occur for $t > 0.5$. This can be understood by recalling how the production rules are chosen. For the $t < 0.5$ regime, rule

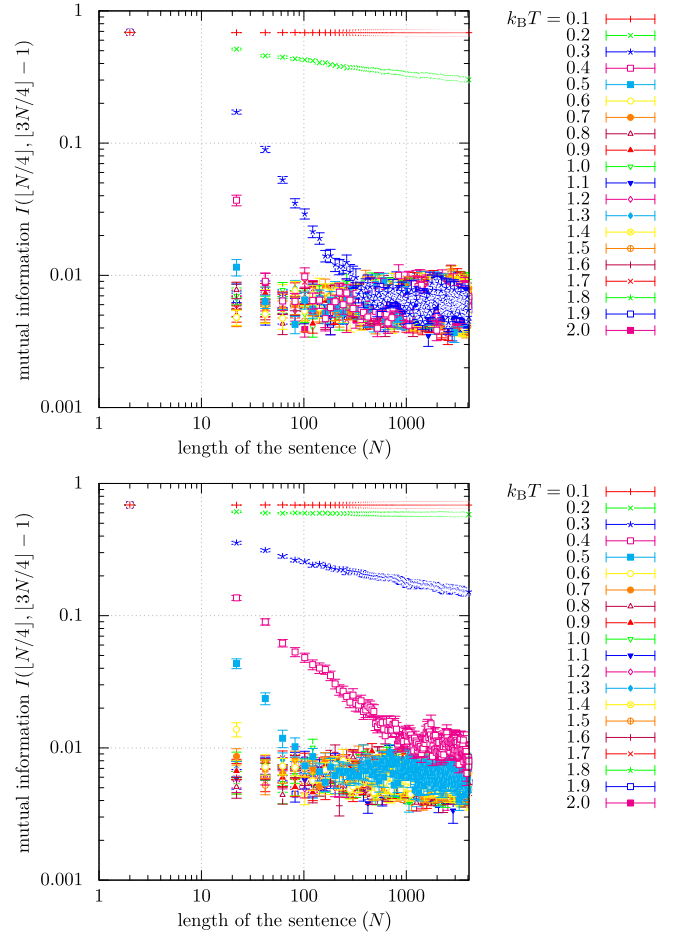


FIG. 19. The mutual information $I(\lfloor N/4 \rfloor, \lfloor 3N/4 \rfloor - 1)$. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 2, J = 1.0, t = 0.0$, and $\epsilon = 0.00$. We show the results for various $k_B T = 0.1, 0.2, \dots, 2.0$, and the curves for different $k_B T$ are overlaid.

$X \rightarrow YZ$ is applied more frequently than $X \rightarrow XX$, so the increasing rate of nonterminal symbols tends to be faster than the speed of decrease of nonterminal symbols. Therefore, each sentence tends to be long and the model behaves like $t = 0$. In contrast, for $t > 0.5$ regime, the decrease dominates over the increase, so short sentences are generated frequently. In our simulations, we set the start symbol of each sentence after the first one to be the last symbol of the previous sentence to introduce correlations among sentences. With this correlation, the order of whole sentences is lost, and phase transitions do not occur. It is nontrivial whether phase transition occurs at $t = 0.5$, but we observe phase transition at $t = 0.5$. Moreover, it is also nontrivial that the critical temperatures decrease as t increases.

VI. CONCLUSIONS

We constructed the context-sensitive grammar model inspired by the one-dimensional short-range Potts model and observed a phase transition in the language generated by this model. Moreover, we confirmed that this transition is consistent with a BKT transition based on the behavior of the Binder parameter and the correlation functions.

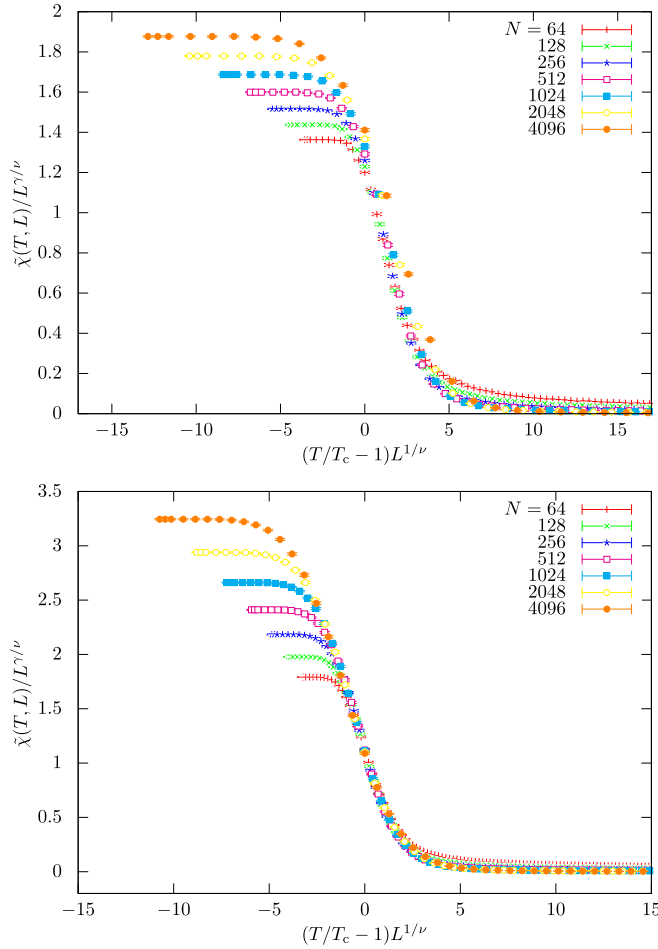


FIG. 20. The finite-size scaling plot. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We plot $\tilde{\chi}(T, N)/N^{\gamma/\nu}$ vs $N^{1/\nu}t$. The critical exponents are set to $\nu = 3.25$ and $\gamma = 3.00$ for $q = 10^{-2.0}$, and to $\nu = 3.50$ and $\gamma = 3.00$ for $q = 10^{-1.0}$. We set $K = 2$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for $N = 64, 128, 256, 512, 1024, 2048, 4096$, and the curves for different N are overlaid.

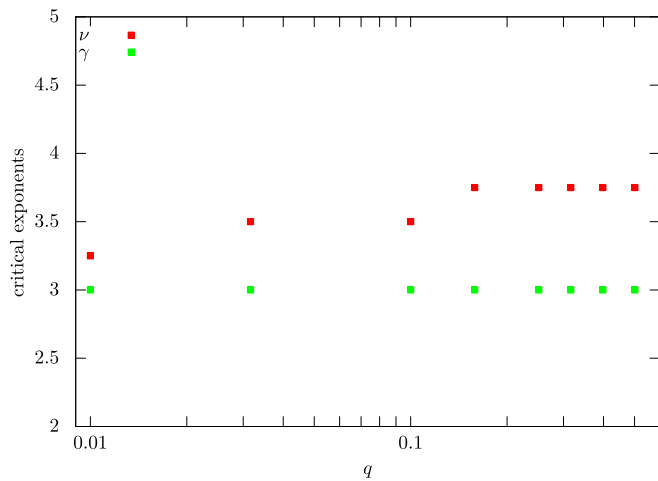


FIG. 21. The q dependence of the critical exponents ν and γ . We set $K = 2$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for $N = 64, 128, 256, 512, 1024, 2048, 4096$, and the curves for different N are overlaid.

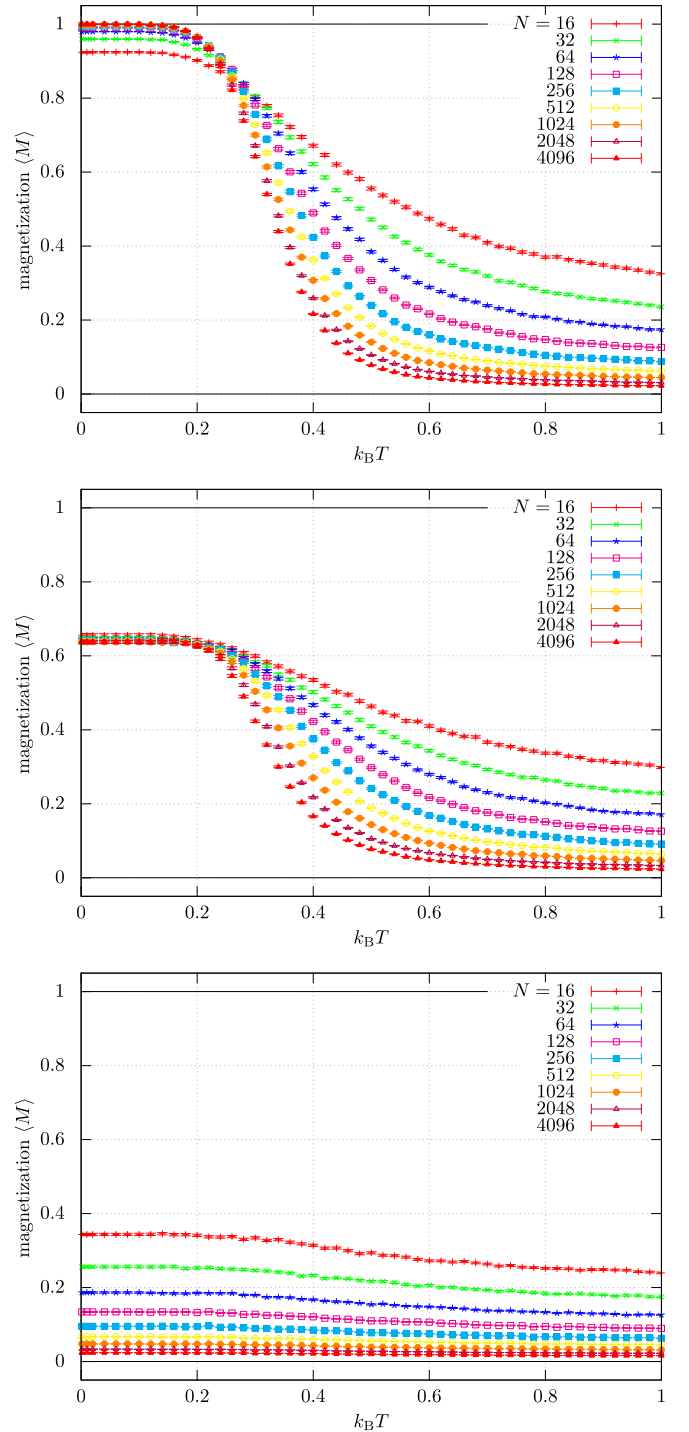


FIG. 22. Temperature dependence of the magnetization [Eq. (3.1)]. The first panel is the result for $t = 0.3$, the second panel is the result for $t = 0.5$, and the third panel is the result for $t = 0.7$. We set $K = 2$, $J = 1.0$, $q = 10^{-2.0}$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

Since it is widely believed that one-dimensional equilibrium systems with short-range interactions do not exhibit phase transitions, the linguistic process investigated in this study appears to induce nonequilibrium effects or effective long-range interactions. Although the present model is highly

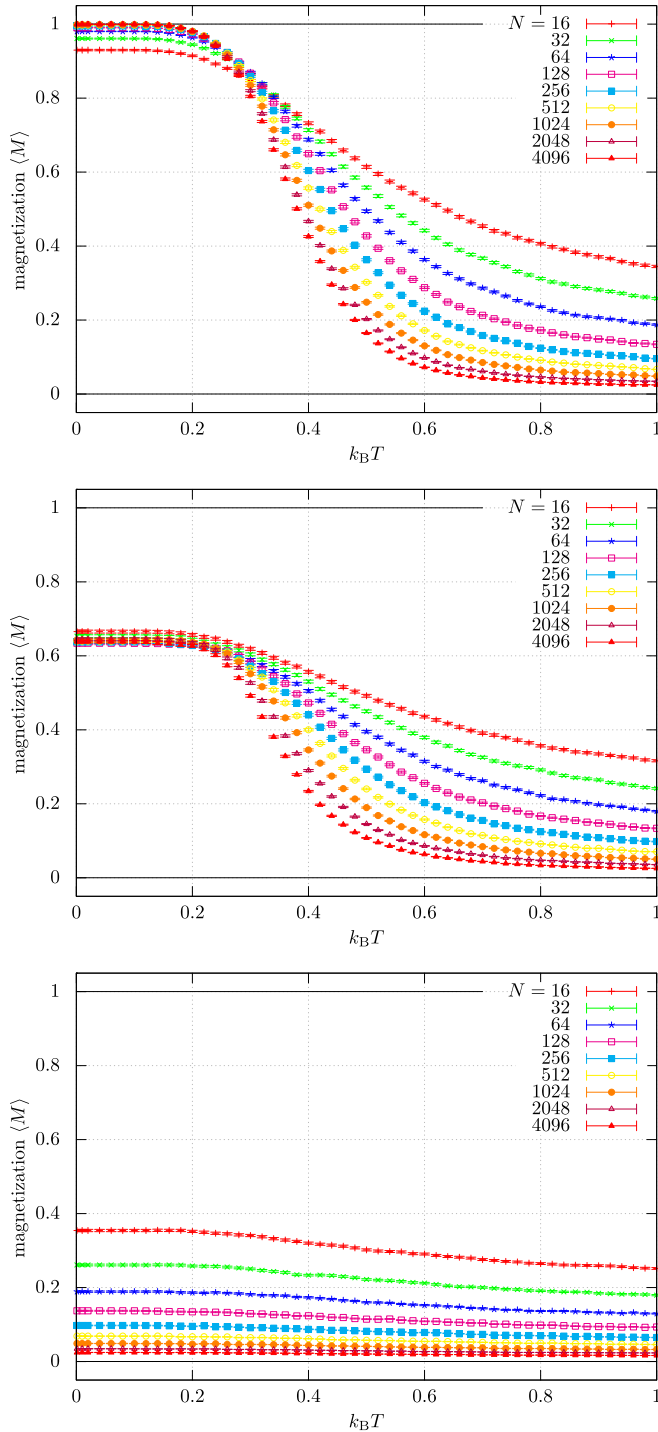


FIG. 23. Temperature dependence of the magnetization [Eq. (3.1)]. The first panel is the result for $t = 0.3$, the second panel is the result for $t = 0.5$, and the third panel is the result for $t = 0.7$. We set $K = 2$, $J = 1.0$, $q = 10^{-1.0}$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

idealized and does not incorporate many important features of natural language, such as semantic constraints, distinctions between syntactic categories and lexical items, or realistic grammar statistics, its purpose is not to faithfully reproduce linguistic data. Rather, the model serves as a minimal theoret-

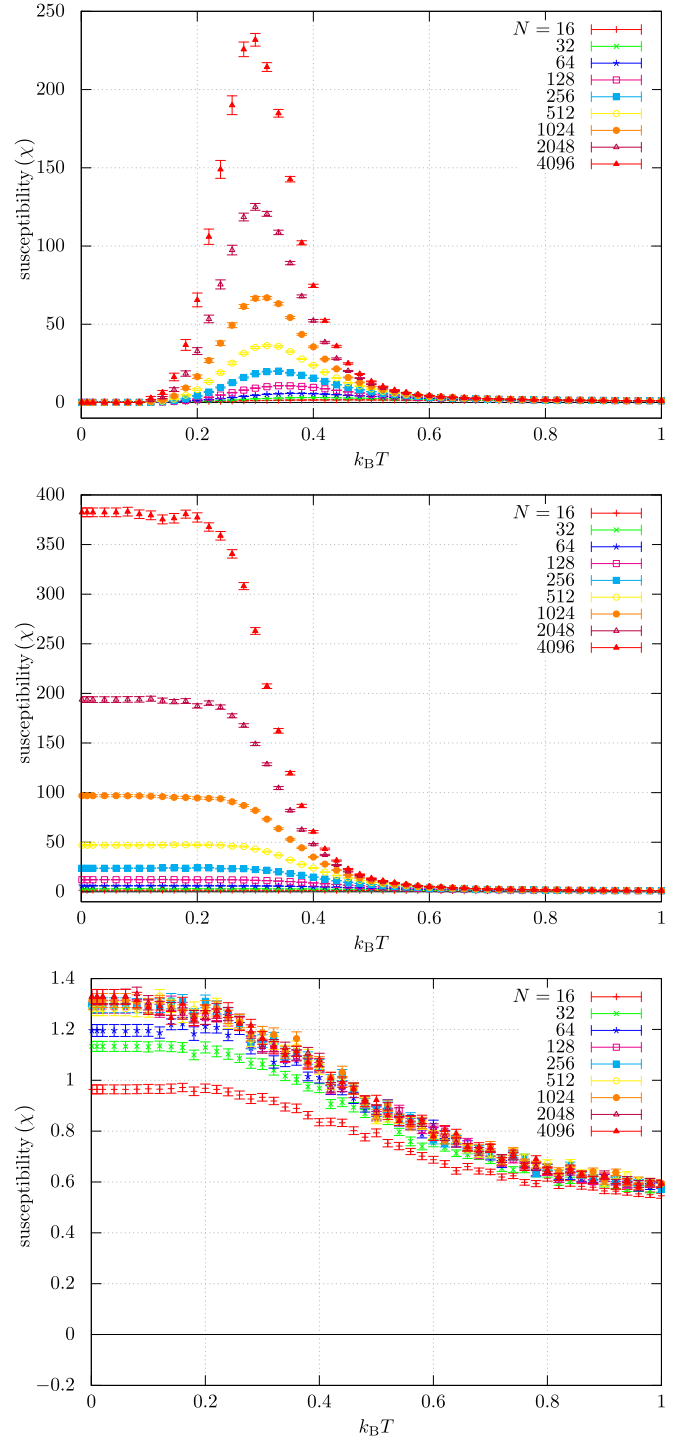


FIG. 24. Temperature dependence of the susceptibility [Eq. (3.4)]. The first panel is the result for $t = 0.3$, the second panel is the result for $t = 0.5$, and the third panel is the result for $t = 0.7$. We set $K = 2$, $J = 1.0$, $q = 10^{-2.0}$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

ical framework for investigating how generative processes can influence collective statistical behavior. The observation that a simple branching process can induce BKT-like behavior and Zipf-like statistics suggests that certain universal properties of

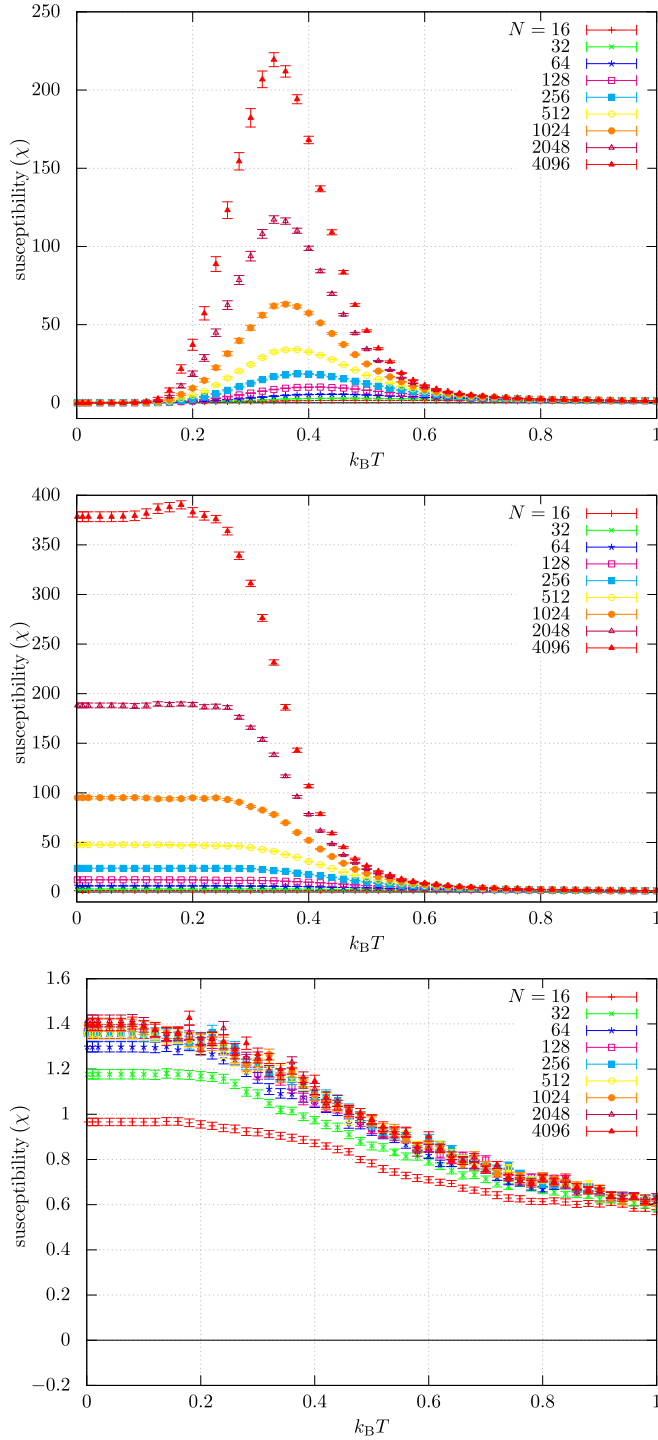


FIG. 25. Temperature dependence of the susceptibility [Eq. (3.4)]. The first panel is the result for $t = 0.3$, the second panel is the result for $t = 0.5$, and the third panel is the result for $t = 0.7$. We set $K = 2$, $J = 1.0$, $q = 10^{-1.0}$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

language may arise from fundamental generative mechanisms rather than from the full complexity of natural language. Extending the present framework to incorporate more realistic linguistic structures remains an important direction for future work.

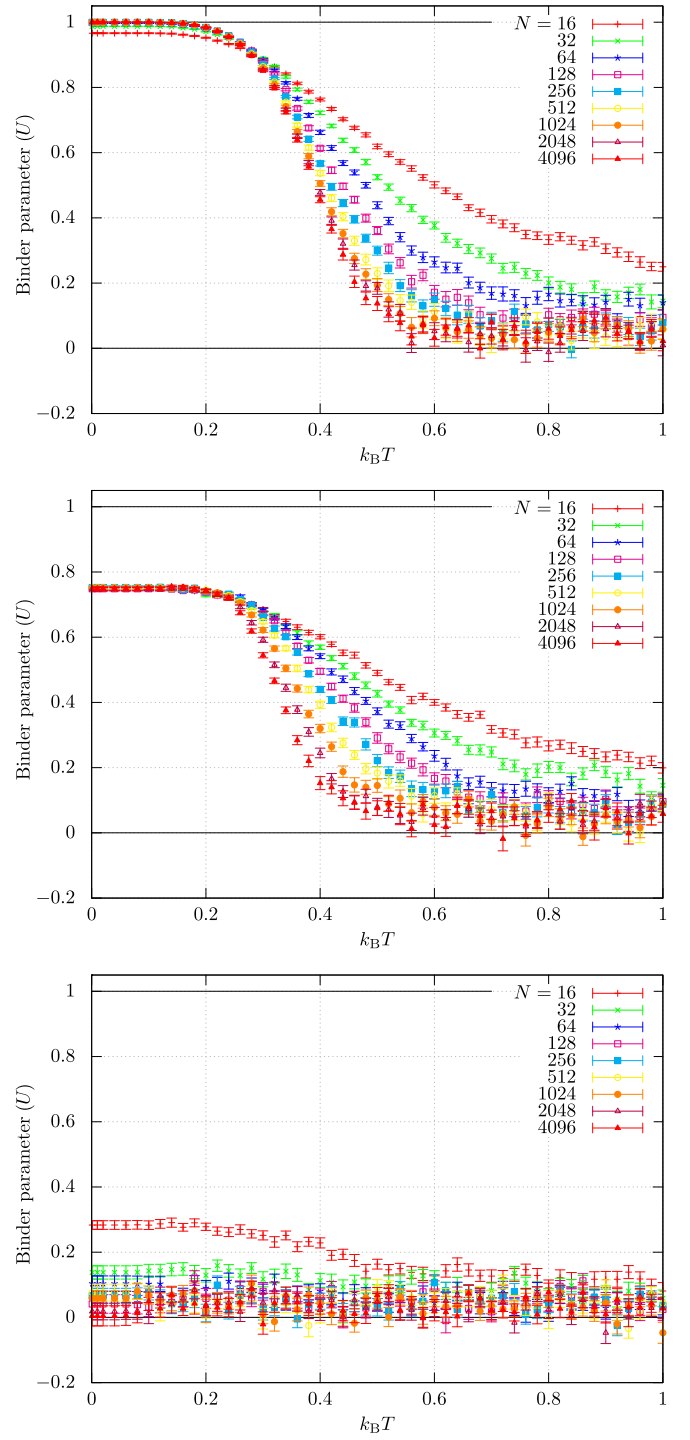


FIG. 26. Temperature dependence of the Binder parameter [Eq. (3.5)]. The first panel is the result for $t = 0.3$, the second panel is the result for $t = 0.5$, and the third panel is the result for $t = 0.7$. We set $K = 2$, $J = 1.0$, $q = 10^{-2.0}$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

ACKNOWLEDGMENT

H.M. was supported by JSPS KAKENHI Grants No. JP25H01499, No. JP26H01783, No. JP26K17043, and No. JPMJBY25E6.

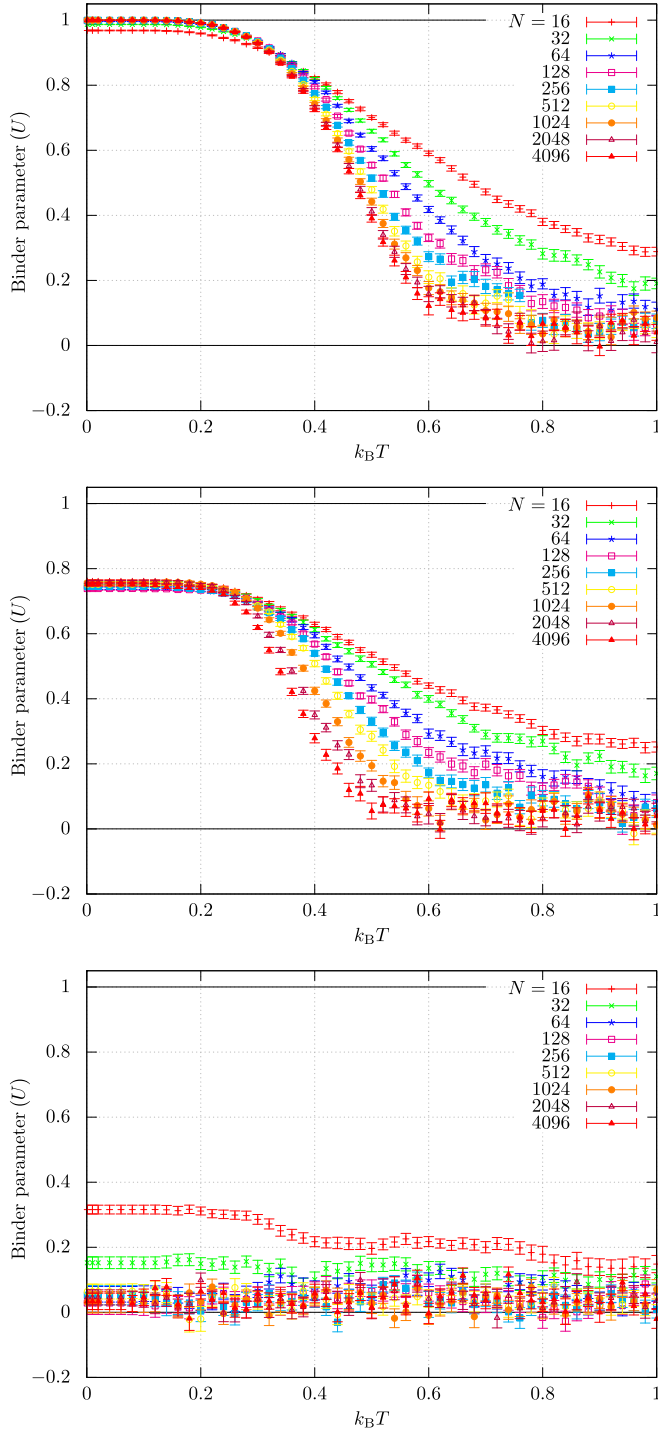


FIG. 27. Temperature dependence of the Binder parameter [Eq. (3.5)]. The first panel is the result for $t = 0.3$, the second panel is the result for $t = 0.5$, and the third panel is the result for $t = 0.7$. We set $K = 2$, $J = 1.0$, $q = 10^{-1.0}$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

DATA AVAILABILITY

We will provide all the numerical codes and plot files upon request.

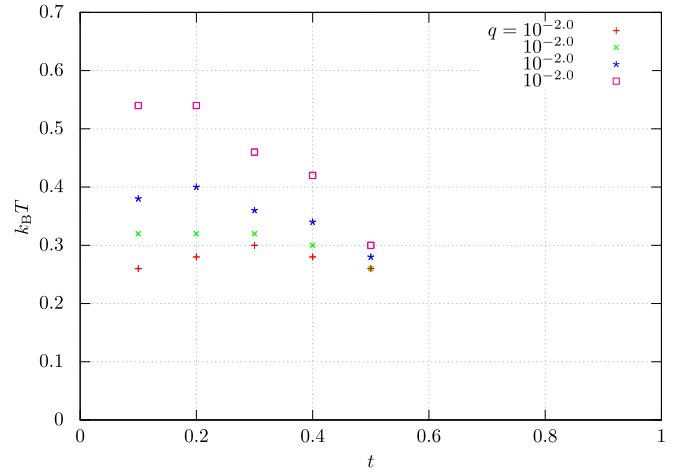


FIG. 28. Phase diagram for $K = 2$, $J = 1.0$, and $\epsilon = 0.00$. The horizontal and vertical axes represent the parameter t , which controls the probability of selecting production rules, and the critical temperature $k_B T$, respectively.

APPENDIX: NUMERICAL SIMULATIONS

We provide additional numerical simulations to support the claims in the main text. In this section, we present a comprehensive analysis across various parameter sets to demonstrate the robustness of the observed phase transition.

Motivated by the finding that mutual information exhibits a power-law scaling in CFGs [31], we also investigated the mutual information of our model in this Appendix. The mutual information $I(i, j)$ between two symbols σ_i and σ_j at positions i and j is defined as

$$I(i, j) := \sum_{\sigma_i, \sigma_j=1}^K p(\sigma_i, \sigma_j) \ln \frac{p(\sigma_i, \sigma_j)}{p(\sigma_i)p(\sigma_j)}, \quad (\text{A1})$$

where $p(\sigma_i)$ and $p(\sigma_j)$ denote the marginal probability distributions of the symbols at positions i and j , respectively, and $p(\sigma_i, \sigma_j)$ represents their joint probability distribution.

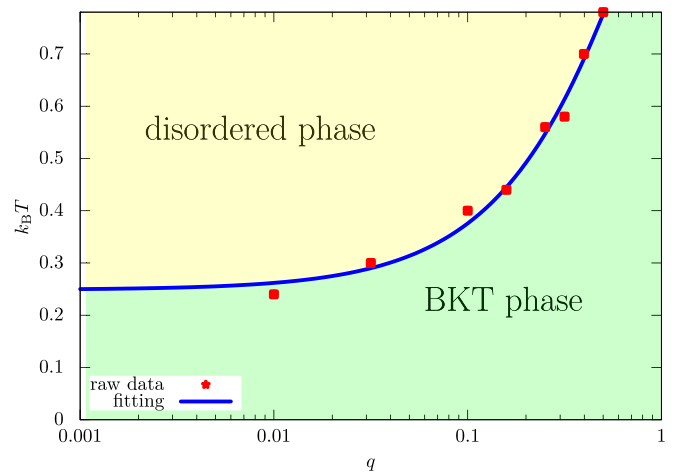


FIG. 29. Phase diagram for $K = 20$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. The horizontal and vertical axes represent the parameter q , which controls the probability of selecting production rules, and the critical temperature $k_B T$, respectively.

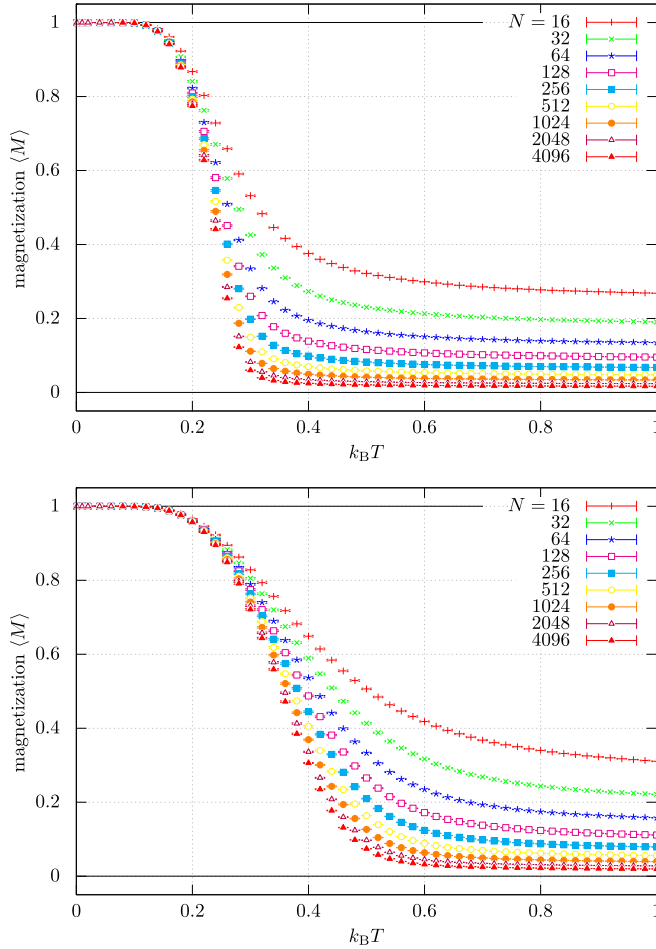


FIG. 30. Temperature dependence of the magnetization [Eq. (3.1)]. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 20$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

To calculate these quantities in our numerical simulations, we estimate these probability distributions using sample frequencies.

1. Case of $K = 2$, $\epsilon = 0.00$, and $t = 0.0$

In particular, we examine the case of $K = 2$ to systematically investigate how the characteristics of the phase transition depend on the number of symbols K . In the framework of statistical mechanics, it is well known that the two-dimensional Potts model exhibits a second-order phase transition when the number of states $K_{\text{Potts}} \leq 4$, whereas it shows a first-order phase transition for $K_{\text{Potts}} > 4$ [20]. Since the number of unique symbols K in our random language model plays a role analogous to the number of states in the Potts model, exploring the numerical behavior at $K = 2$ allows us to test whether the system's criticality is robust across these traditionally distinct regimes.

a. $k_B T$ - q phase diagram

In Fig. 8, we show a $k_B T$ - q phase diagram of $K = 2$ and $\epsilon = 0.00$. As shown in the phase diagram, the overall behavior

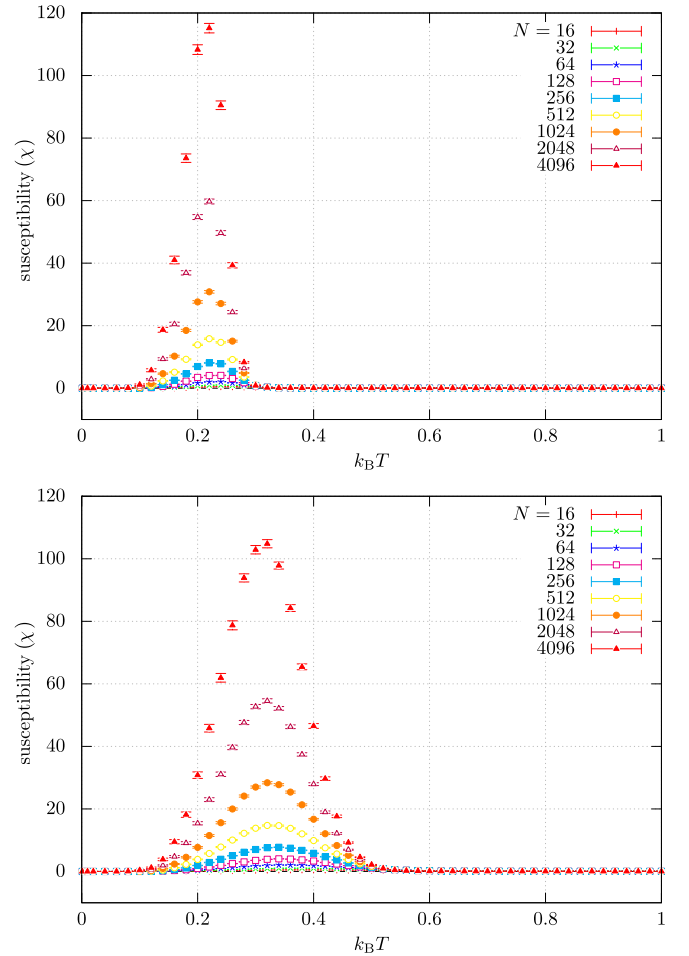


FIG. 31. Temperature dependence of the susceptibility [Eq. (3.4)]. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 20$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

of the phase boundary is qualitatively similar to that observed in the $K = 20$ case.

b. Magnetization, susceptibility, and Binder parameter

Below, we compare the results for $q = 10^{-2.0}$ and $q = 10^{-1.0}$. The transition temperatures are estimated to be $k_B T = 0.20$ for $q = 10^{-2.0}$ and $k_B T = 0.34$ for $q = 10^{-1.0}$. The magnetization is shown in Fig. 9, the susceptibility is shown in Fig. 10, and the Binder parameter is shown in Fig. 11. The magnetization and susceptibility generally follow the same trends as in the $K = 20$ case. However, the Binder parameter does not exhibit a dip near the transition temperature. The absence of a negative dip can be interpreted as the absence of phase coexistence, characterized by a single-peaked magnetization distribution.

We also show the system-size dependence of the magnetization, susceptibility, and Binder parameter. Figure 12 shows the system-size dependence of the magnetization, Fig. 13 shows that of the susceptibility, and Fig. 14 shows that of the Binder parameter. It should be noted that whether the

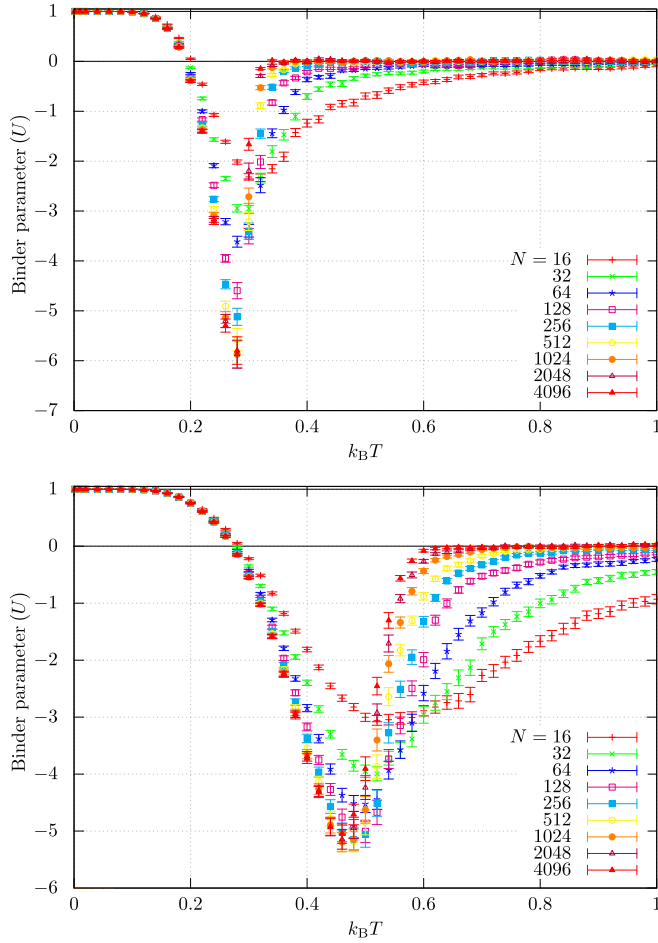


FIG. 32. Temperature dependence of the Binder parameter [Eq. (3.5)]. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 20$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

system-size dependence of the susceptibility grows linearly or exhibits curvature in the log-log plot is a crucial factor for accurately estimating the transition temperature.

c. Histogram of the magnetization

The histogram of the magnetization is shown in Fig. 15. It shows a sharp peak near $\langle M \rangle = 1$ in the low-temperature regime and near $\langle M \rangle = 0$ in the high-temperature regime. This variation is highly consistent with the standard behavior expected in typical phase transitions. As discussed in the context of the Binder parameter (Fig. 11), no double-peak structure is observed.

d. Typical configurations of symbols

By examining the histograms shown in Appendix A 1 c, we can identify, for each temperature, the most frequent value of the magnetization. We regard this value as the typical magnetization at that temperature, and we show, as examples, the configurations of the generated text whose magnetization takes this typical value in Figs. 16 and 17. The temperatures

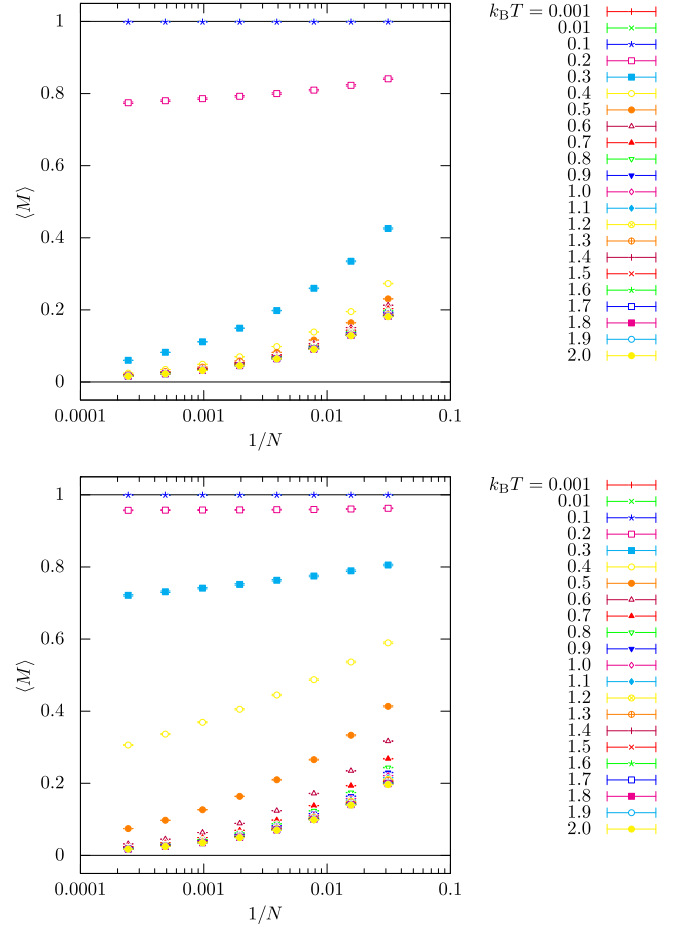


FIG. 33. The system-size dependence of the magnetization [Eq. (3.1)]. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 20$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various $k_B T = 0.001, 0.01, 0.1, 0.2, \dots, 2.0$, and the curves for different $k_B T$ are overlaid.

shown correspond to values below the transition temperature, just above the transition temperature, and above the transition temperature. For both $q = 10^{-2.0}$ and $10^{-1.0}$, we can visually confirm the microscopic changes in the generated sentences. As the temperature increases across the transition point, the generated sequences clearly transition from an ordered state, where identical symbols are aligned, to a disordered state characterized by randomly mixed symbols.

e. Correlation function

The data of the correlation function $\tilde{G}([N/4], [3N/4] - 1)$ are shown in Fig. 18. Below the transition temperature, the correlation function exhibits a power-law decay, as evidenced by the linear behavior in the log-log plot, suggesting the presence of a critical phase.

f. Mutual information

The data of the mutual information $I([N/4], [3N/4] - 1)$ are shown in Fig. 19. Similar to the correlation function, the mutual information displays a power-law decay below the transition temperature.

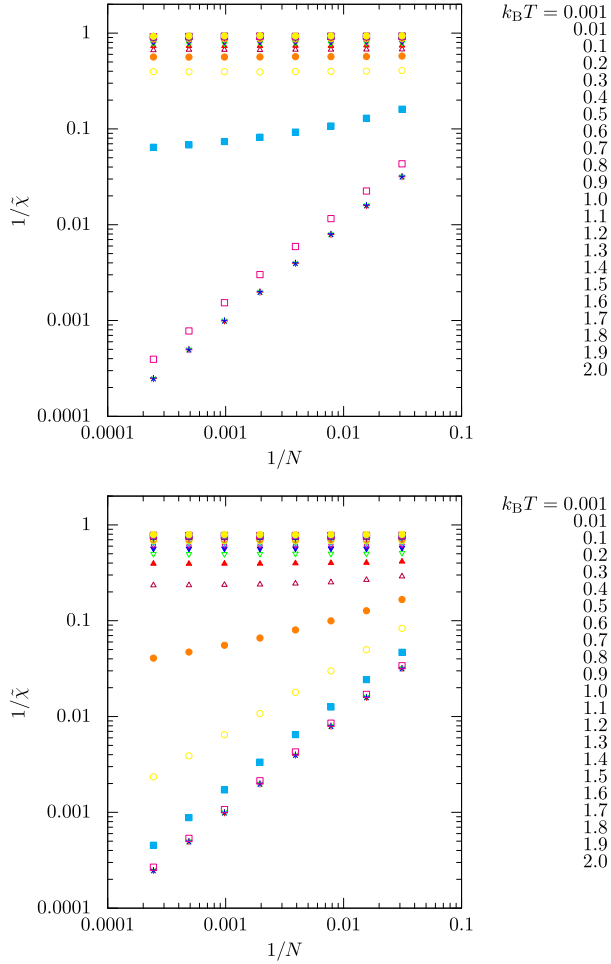


FIG. 34. The system-size dependence of the susceptibility [Eq. (3.4)]. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 20$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various $k_B T = 0.001, 0.01, 0.1, 0.2, \dots, 2.0$, and the curves for different $k_B T$ are overlaid.

g. Finite-size scaling

The finite-size scaling plot is shown in Fig. 20. The scaled data points collapse well onto a single curve near the transition temperature, allowing us to determine the critical exponents by evaluating the quality of the data collapse.

The q dependence of the critical exponents γ and ν is shown in Fig. 21. The tendency that the critical exponent γ remains roughly constant, while ν increases with respect to q is consistent with the behavior previously observed in the random language model with long-range interactions.

2. Case of $K = 2$, $\epsilon = 0.00$, and $t \neq 0$

When $t \neq 0$, terminal symbols can be generated during the sentence production process [rule (2.1a)]. For $t > 0.5$, the probability of generating a terminal symbol exceeds the probability of symbol branching. As a result, the sentence generation process terminates before the sentence length becomes sufficiently large, immediately moving to the next sentence generation. This rapid termination prevents symbols from es-

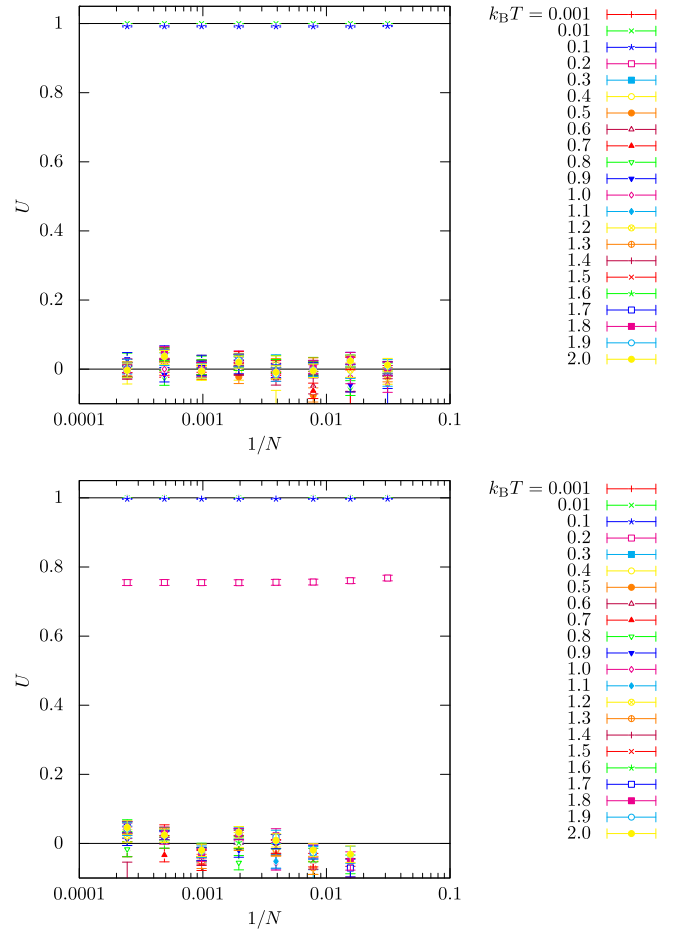


FIG. 35. The system-size dependence of the Binder parameter [Eq. (3.5)]. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 20$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various $k_B T = 0.001, 0.01, 0.1, 0.2, \dots, 2.0$, and the curves for different $k_B T$ are overlaid.

tablishing sufficient correlations, leading to the disappearance of the phase transition. On the other hand, at $t = 0.5$, we observe a phenomenon where the magnetization does not develop long-range order but instead saturates at a certain constant value. The underlying mechanism behind this saturation behavior remains unclarified and is left as an intriguing open question for future investigation.

a. Magnetization, susceptibility, and Binder parameter

Below, we compare the results for $t = 0.3, 0.5, 0.7$ and for $q = 10^{-2.0}, 10^{-1.0}$. The transition temperatures are estimated to be those values shown in Table I.

The magnetization is shown in Figs. 22 and 23, the susceptibility is shown in Figs. 24 and 25, and the Binder parameter is shown in Figs. 26 and 27. As can be seen from the behavior of the magnetization, susceptibility, and Binder parameter, the phase transition completely disappears at $t = 0.7$. Conversely, at $t = 0.5$, we observe that in the low-temperature regime, the magnetization saturates at approximately 0.65 for both $q = 10^{-2.0}$ and $10^{-1.0}$, and the Binder parameter similarly saturates around 0.75. The physical origin of this saturation

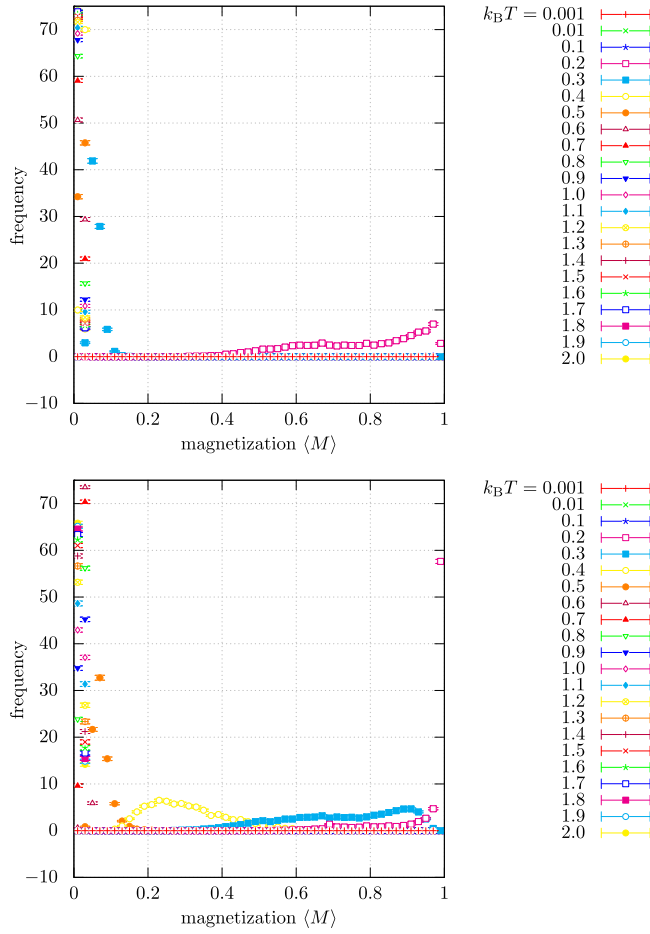


FIG. 36. The histogram of the magnetization. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 20$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various $k_B T = 0.001, 0.1, \dots, 2.0$, and the curves for different $k_B T$ are overlaid.

behavior has not yet been clarified and remains an open question for future investigation.

b. $k_B T$ - t phase diagram

The $k_B T$ - t phase diagram is shown in Fig. 28. In the $k_B T$ - t phase diagram, we observe a nontrivial behavior where the transition temperatures for different values of q show a tendency to converge toward a certain finite value (around 0.27) as t approaches 0.5.

3. Case of $K = 20$, $\epsilon = 0.00$, and $t = 0.0$

Although many of the figures for $q = 10^{-2.0}$ have already been discussed in the main text, they are reproduced here to facilitate a direct comparison across with those for $q = 10^{-1.0}$.

a. $k_B T$ - q phase diagram

In Fig. 29, we show the $k_B T$ - q phase diagram of $K = 20$ and $\epsilon = 0.00$. The phase diagram presented here is identical to the one discussed in the main text (Fig. 6), provided again to serve as a baseline for the following detailed analyses.

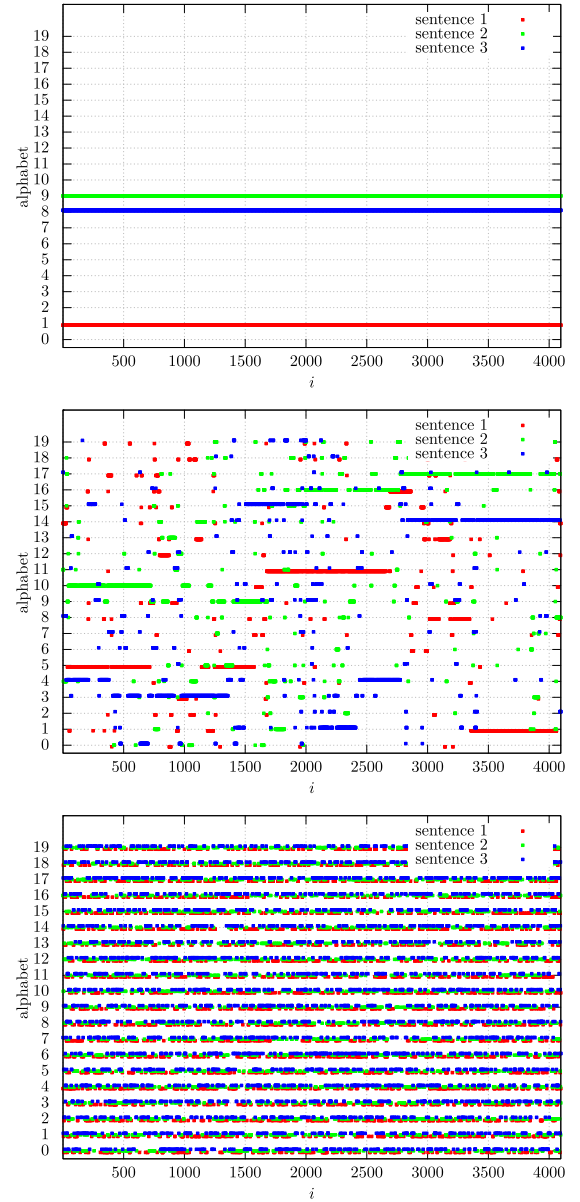


FIG. 37. Typical configurations of symbols. The upper panel shows an example sentence at a temperature below the transition ($k_B T = 0.001$) with magnetization $0.98 \leq M < 1.00$. The middle panel shows an example sentence at a temperature just above the transition ($k_B T = 0.20$) with magnetization $0.98 \leq M < 1.00$. The lower panel shows an example sentence at a temperature above the transition ($k_B T = 0.40$) with magnetization $0.00 \leq M < 0.02$. We set $K = 20$, $J = 1.0$, $q = 10^{-2.0}$, $t = 0.0$, and $\epsilon = 0.00$.

b. Magnetization, susceptibility, and Binder parameter

Below, we compare the results for $q = 10^{-2.0}$, $10^{-1.0}$. The transition temperatures are estimated to be $k_B T = 0.20$ for $q = 10^{-2.0}$ and $k_B T = 0.34$ for $q = 10^{-1.0}$. The magnetization is shown in Fig. 30, the susceptibility is shown in Fig. 31, and the Binder parameter is shown in Fig. 32. Unlike the case of $K = 2$, the Binder parameter for $K = 20$ exhibits a distinct dip near the transition temperature, suggesting the existence of phase coexistence.

We also show the system-size dependence of the magnetization, susceptibility, and Binder parameter. Figure 33 shows

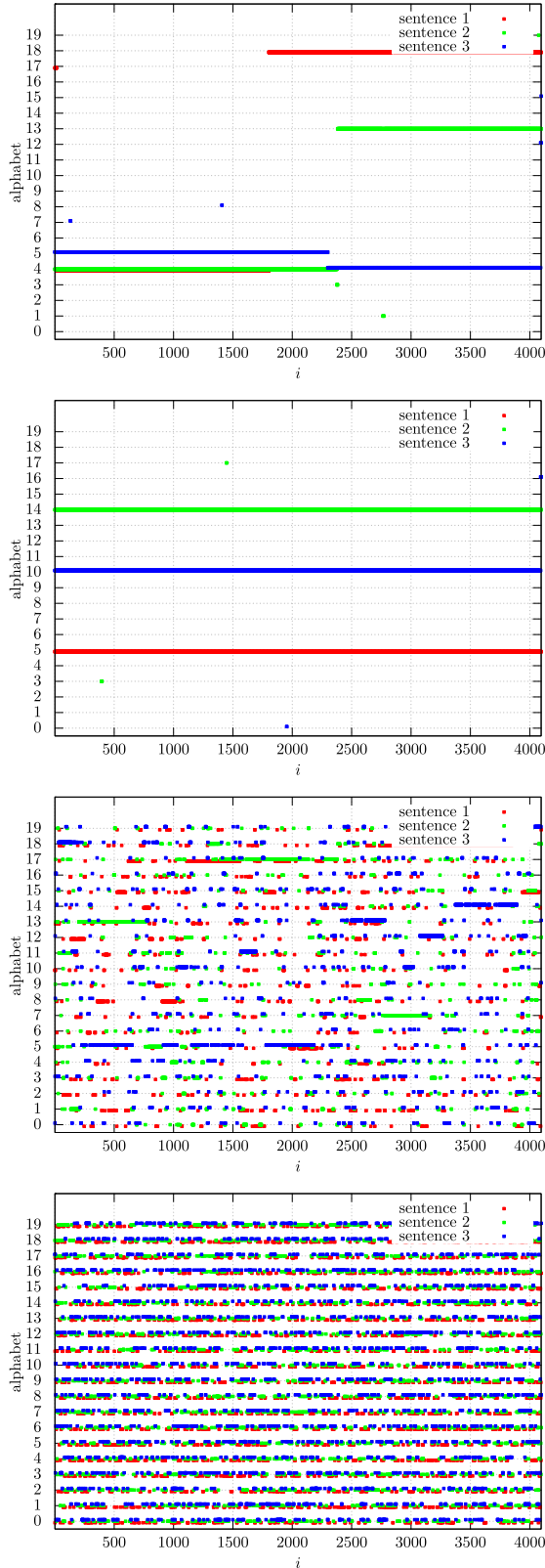


FIG. 38. Typical configurations of symbols. The first panel shows an example sentence at a temperature below the transition ($k_B T = 0.20$) with magnetization $0.68 \leq M < 0.70$. The second panel shows an example sentence at the same temperature ($k_B T = 0.20$) with magnetization $0.98 \leq M < 1.00$. The third panel shows an example sentence at a temperature just above the transition ($k_B T = 0.40$) with magnetization $0.22 \leq M < 0.24$. The

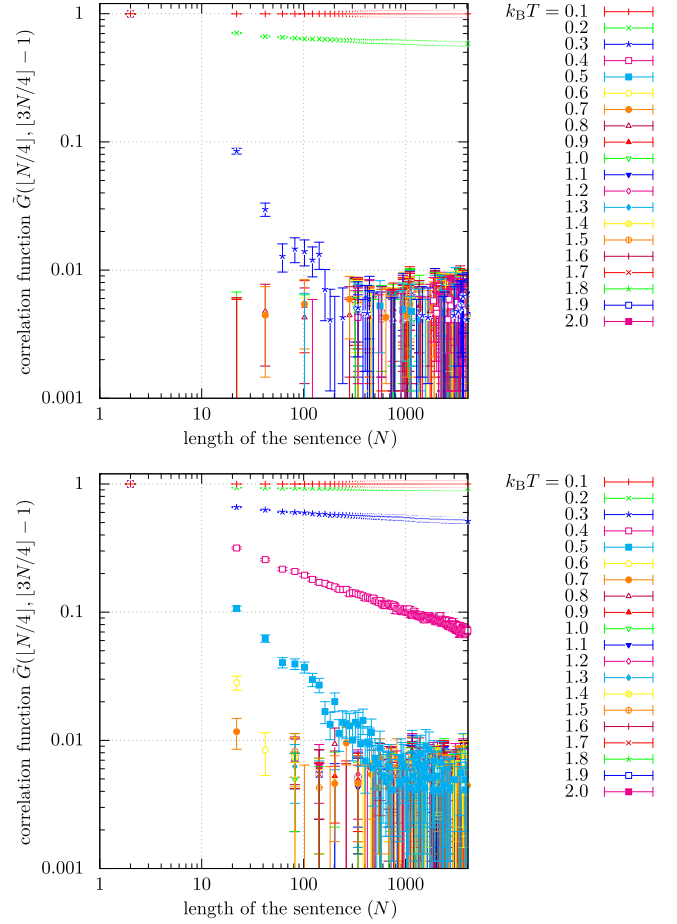


FIG. 39. The correlation function [Eq. (3.7)], with $i = \lfloor N/4 \rfloor$ and $j = \lfloor 3N/4 \rfloor - 1$. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We set $K = 20$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for various $k_B T = 0.1, 0.2, \dots, 2.0$, and the curves for different $k_B T$ are overlaid.

the system-size dependence of the magnetization, Fig. 34 shows that of the susceptibility, and Fig. 35 shows that of the Binder parameter.

c. Histogram of the magnetization

The histogram of the magnetization is shown in Fig. 36. Intriguingly, the probability distribution of the magnetization for $q = 10^{-1.0}$ at $k_B T = 0.20$ clearly displays a bimodal structure with two coexisting peaks. This bimodal behavior is not observed in the $K = 2$ case and is highly suggestive of a phase coexistence. This result qualitatively reflects the properties of the standard Potts model, which undergoes a first-order transition for larger state numbers ($K > 4$).

fourth panel shows an example sentence at a temperature above the transition ($k_B T = 0.60$) with magnetization $0.02 \leq M < 0.04$. We set $K = 20$, $R_- = R_+ = 1$, $J = 0$, $q = 10^{-1.0}$, $t = 0.0$, $s = 0.9$, and $\epsilon = 0.00$. The first and second panels correspond to the same value of $k_B T$, because the histogram at this temperature exhibits two peaks, and we show configurations corresponding to both peaks. However, the configurations with $0.98 \leq M < 1.00$ appear more frequently than those with $0.68 \leq M < 0.70$.

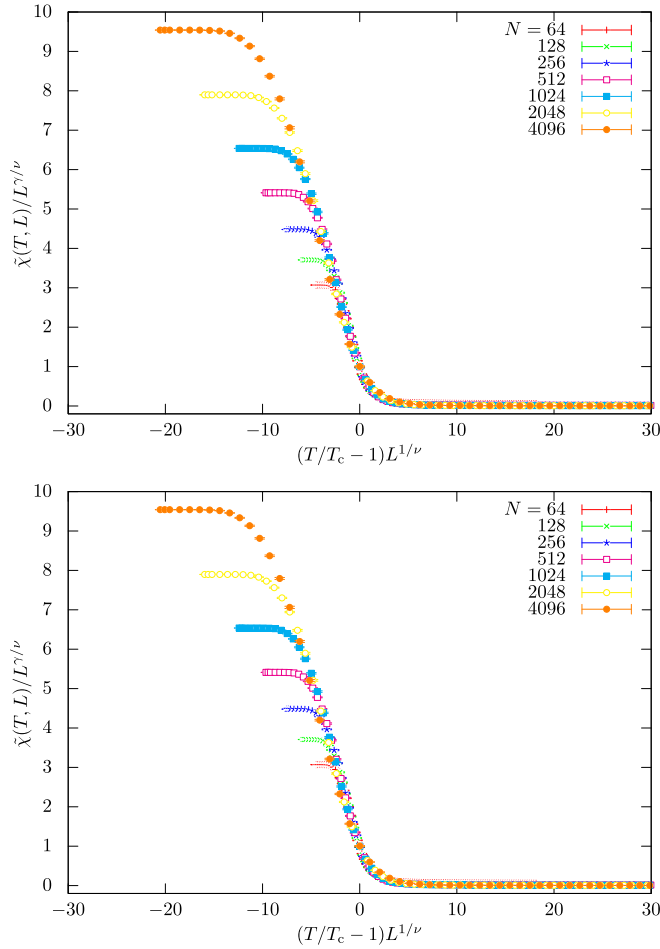


FIG. 40. The finite-size scaling plot. The top panel is the result for $q = 10^{-2.0}$, and the bottom panel is the result for $q = 10^{-1.0}$. We plot $\tilde{\chi}(T, N)/N^{\gamma/\nu}$ vs $N^{1/\nu}t$. The critical exponents are set to $\nu = 2.50$ and $\gamma = 2.00$ for $q = 10^{-2.0}$, and to $\nu = 2.75$ and $\gamma = 2.00$ for $q = 10^{-1.0}$. We set $K = 20$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for $N = 64, 128, 256, 512, 1024, 2048, 4096$, and the curves for different N are overlaid.

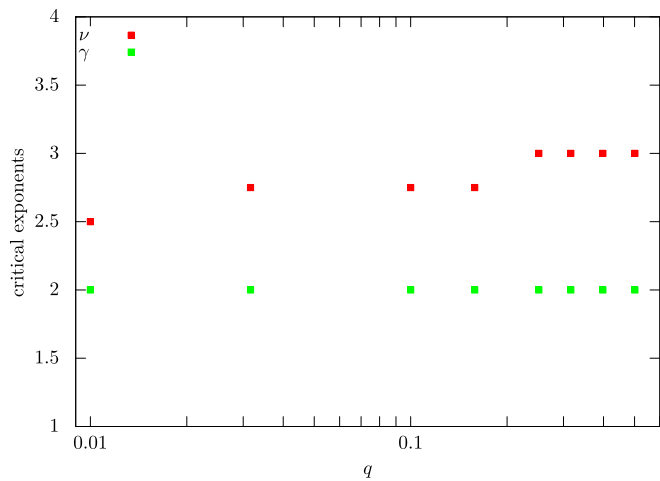


FIG. 41. The q dependence of the critical exponents ν and γ . We set $K = 20$, $J = 1.0$, $t = 0.0$, and $\epsilon = 0.00$. We show the results for $N = 64, 128, 256, 512, 1024, 2048, 4096$, and the curves for different N are overlaid.

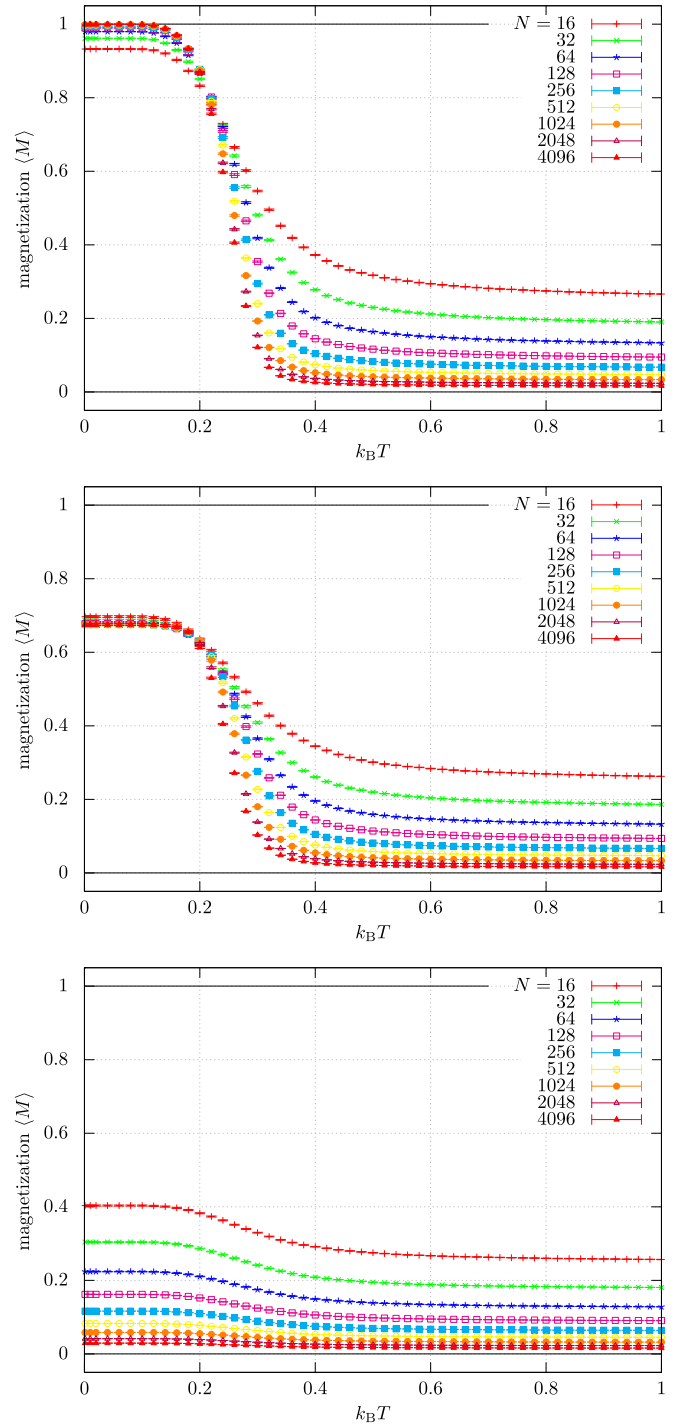


FIG. 42. Temperature dependence of the magnetization [Eq. (3.1)]. The first panel is the result for $t = 0.3$, the second panel is the result for $t = 0.5$, and the third panel is the result for $t = 0.7$. We set $K = 20$, $J = 1.0$, $q = 10^{-2.0}$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

d. Typical configurations of symbols

By examining the histograms shown in Appendix A 3 c, we can identify, for each temperature, the most frequent value of the magnetization. We regard this value as the typical magnetization at that temperature, and we show, as examples,

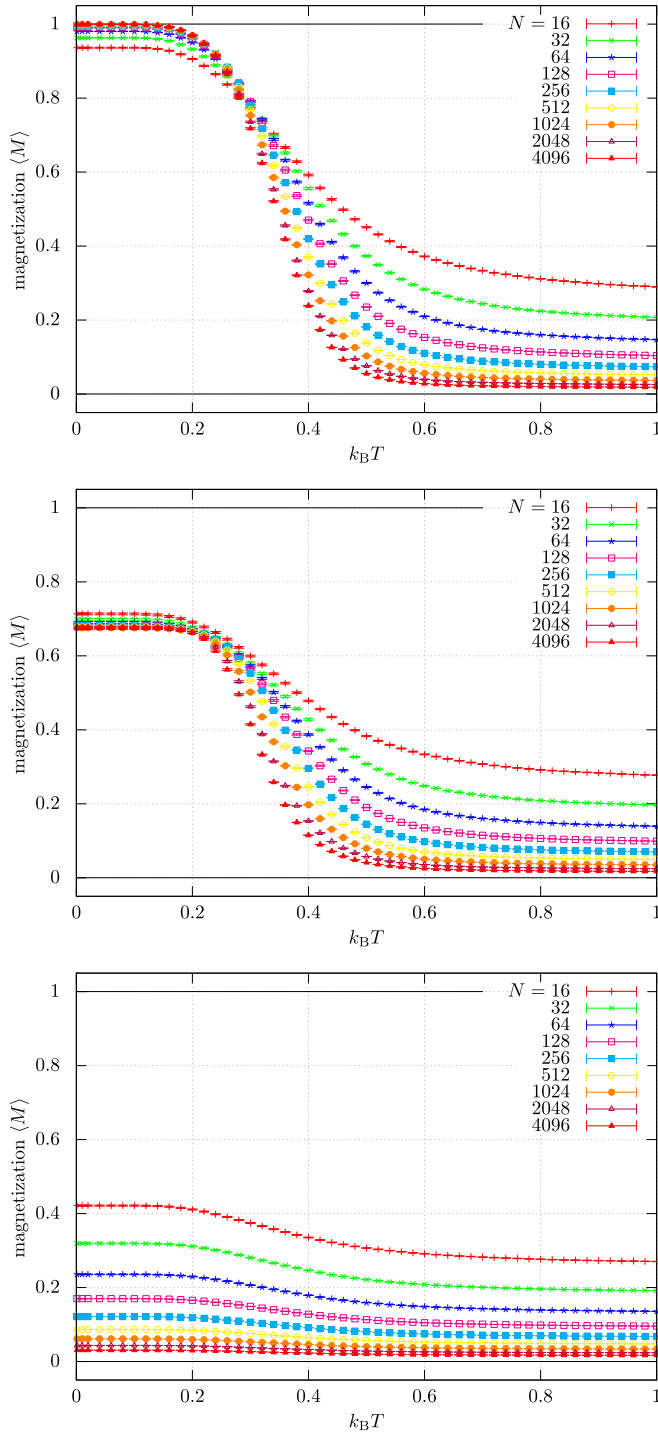


FIG. 43. Temperature dependence of the magnetization [Eq. (3.1)]. The first panel is the result for $t = 0.3$, the second panel is the result for $t = 0.5$, and the third panel is the result for $t = 0.7$. We set $K = 20$, $J = 1.0$, $q = 10^{-1.0}$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

the configurations of the generated text whose magnetization takes this typical value in Figs. 37 and 38. The temperatures shown correspond to values below the transition temperature, just above the transition temperature, and above the transition temperature. For both $q = 10^{-2.0}$ and $10^{-1.0}$, we can visually

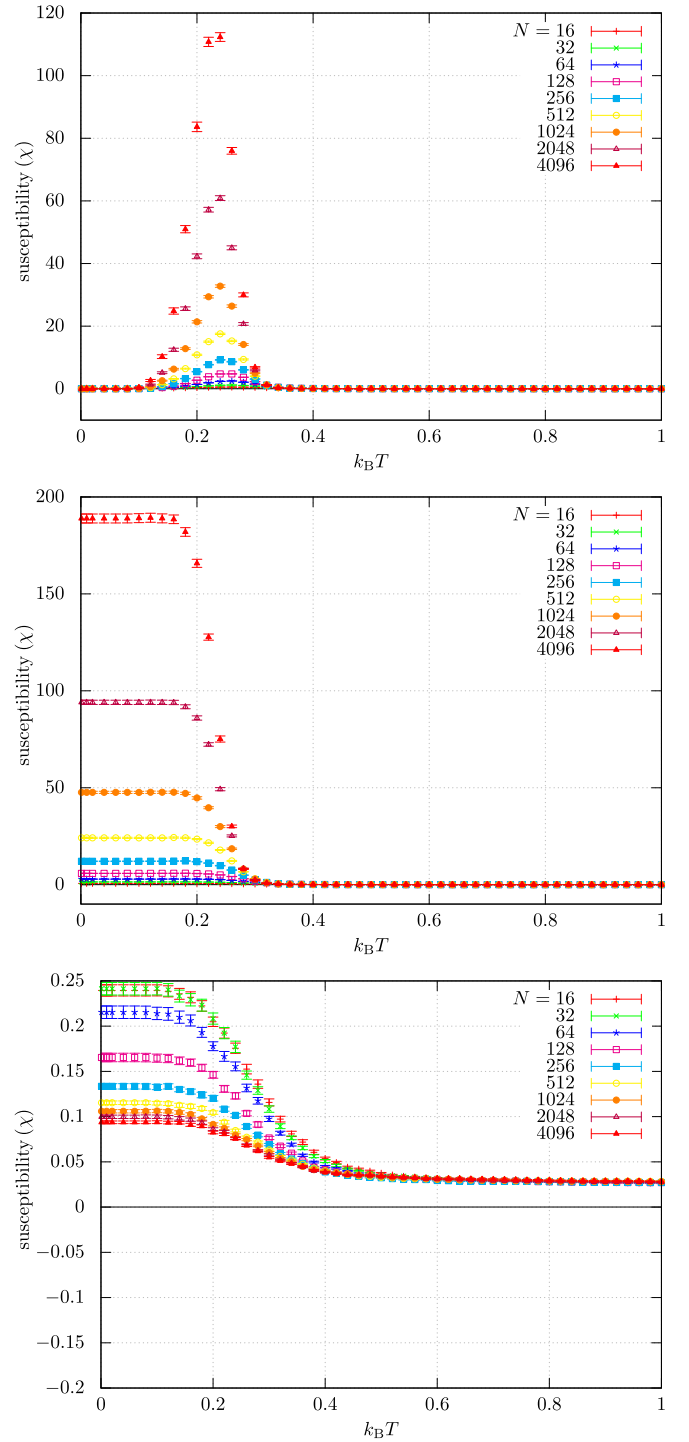


FIG. 44. Temperature dependence of the susceptibility [Eq. (3.4)]. The first panel is the result for $t = 0.3$, the second panel is the result for $t = 0.5$, and the third panel is the result for $t = 0.7$. We set $K = 20$, $J = 1.0$, $q = 10^{-2.0}$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

confirm that as the temperature increases beyond the transition point, the generated sentences change from an ordered configuration of aligned symbols to a disordered, chaotic mixture. In particular, for $q = 10^{-1.0}$ at $k_B T = 0.20$, the configura-

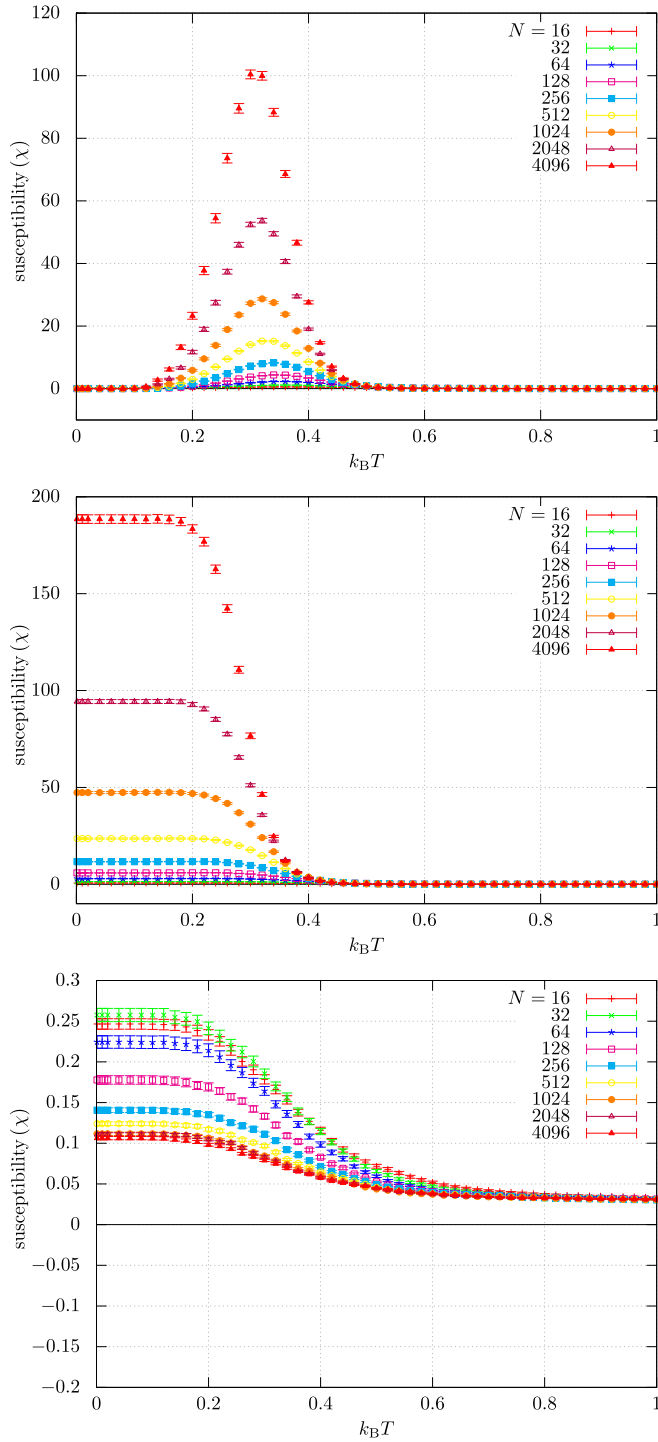


FIG. 45. Temperature dependence of the susceptibility [Eq. (3.4)]. The first panel is the result for $t = 0.3$, the second panel is the result for $t = 0.5$, and the third panel is the result for $t = 0.7$. We set $K = 20$, $J = 1.0$, $q = 10^{-1.0}$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

tions corresponding to the two distinct peaks in the bimodal magnetization histogram are presented. Specifically, when the magnetization lies in the range $0.98 \leq M < 1.00$ (the second panel from the top in Fig. 38), the generated sentence is

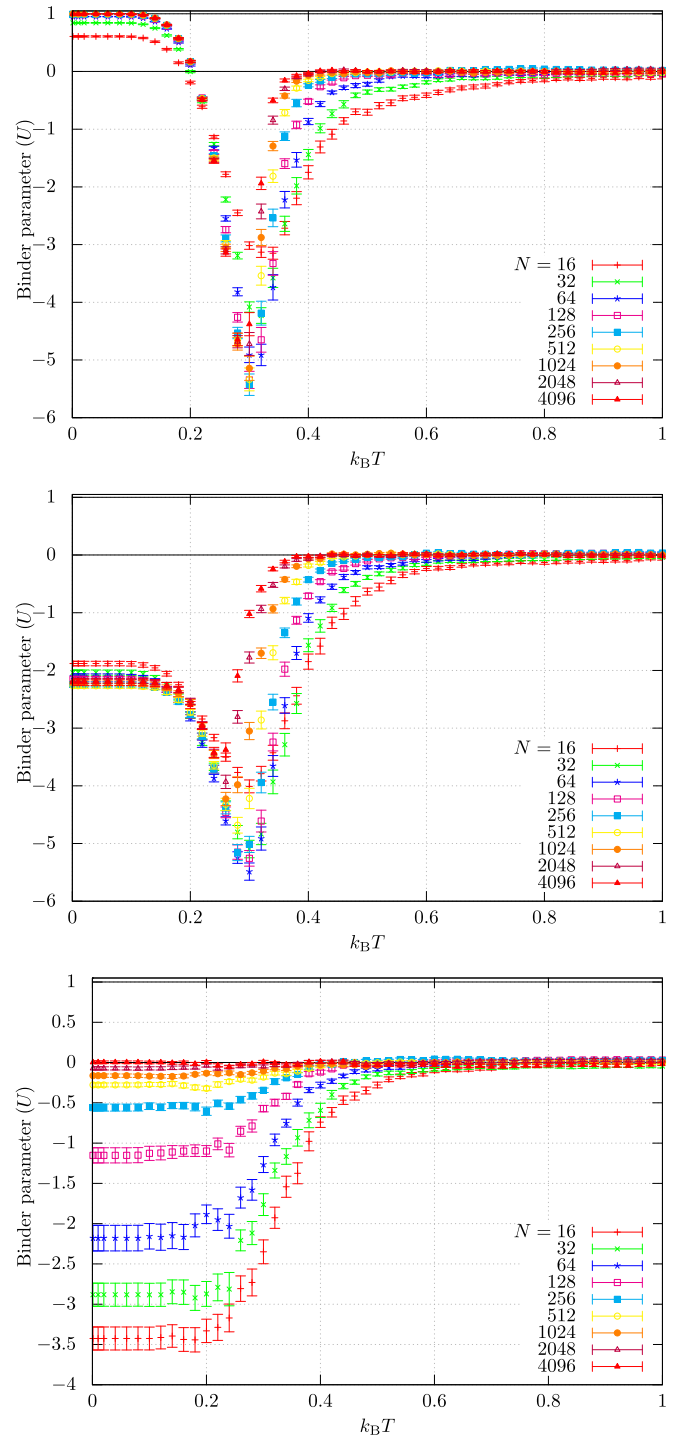


FIG. 46. Temperature dependence of the Binder parameter [Eq. (3.5)]. The first panel is the result for $t = 0.3$, the second panel is the result for $t = 0.5$, and the third panel is the result for $t = 0.7$. We set $K = 20$, $J = 1.0$, $q = 10^{-2.0}$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

ordered and aligned into a single symbol. In contrast, for $0.68 \leq M < 0.70$ (the top panel in Fig. 38), the configuration is visibly segregated into regions aligned with two distinct symbols.

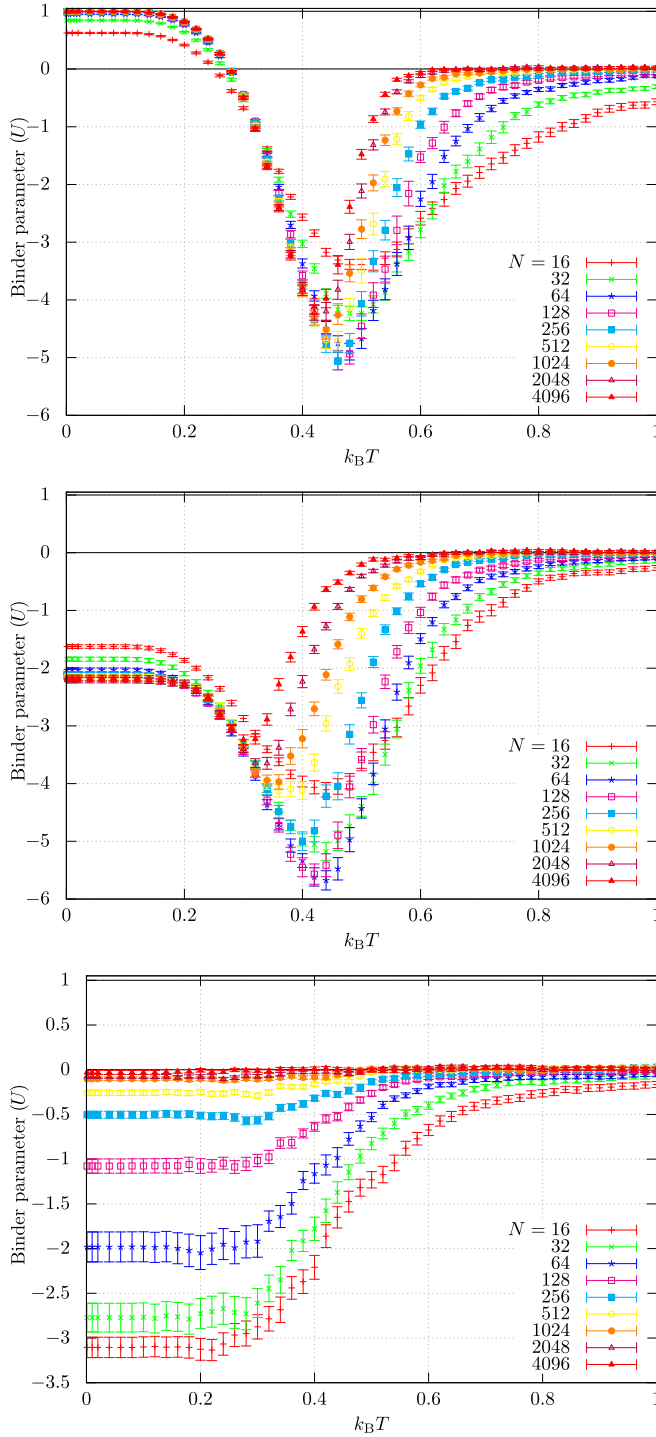


FIG. 47. Temperature dependence of the Binder parameter [Eq. (3.5)]. The first panel is the result for $t = 0.3$, the second panel is the result for $t = 0.5$, and the third panel is the result for $t = 0.7$. We set $K = 20$, $J = 1.0$, $q = 10^{-1.0}$, and $\epsilon = 0.00$. We show the results for various system sizes $N = 16, 32, \dots, 4096$, and the curves for different N are overlaid in each panel.

e. Correlation function

The data of the correlation function $\tilde{G}([N/4], [3N/4] - 1)$ are shown in Fig. 39. Below the transition temperature, the correlation function exhibits a clear power-law decay that is consistent with the BKT mechanism.

TABLE II. The critical temperature of each parameter set. The symbol “—” means no phase transitions occur in the parameter setting.

	$t = 0.3$	$t = 0.5$	$t = 0.7$
$q = 10^{-2.0}$	0.24	0.18	—
$q = 10^{-1.0}$	0.30	0.24	—

f. Finite-size scaling

The finite-size scaling plot is shown in Fig. 40. The scaled data points demonstrate an excellent collapse onto a single curve near the transition temperature.

The q dependence of the critical exponents γ and ν is shown in Fig. 41. Regarding the q -dependence of the critical exponents, we observe that the exponent γ remains approximately constant, while ν systematically increases with q . This trend is qualitatively identical to the behavior found in the $K = 2$ case, suggesting a shared underlying scaling mechanism despite the different transition orders.

4. Case of $K = 20$, $\epsilon = 0.00$, and $t \neq 0$

a. Magnetization, susceptibility, and Binder parameter

Below, we compare the results for $t = 0.3, 0.5, 0.7$ and for $q = 10^{-2.0}, 10^{-1.0}$. The transition temperatures are estimated to be those values shown in Table II.

The magnetization is shown in Figs. 42 and 43, the susceptibility is shown in Figs. 44 and 45, and the Binder parameter is shown in Figs. 46 and 47. As evidenced by the behavior of the magnetization, susceptibility, and Binder parameter, the phase transition is completely suppressed at $t = 0.7$. Conversely, at $t = 0.5$, both $q = 10^{-2.0}$ and $10^{-1.0}$ exhibit a low-temperature saturation of the magnetization at approximately 0.7, accompanied by a corresponding plateaus in the Binder parameter around -2 . This behavior is similar to that for $K = 2$.

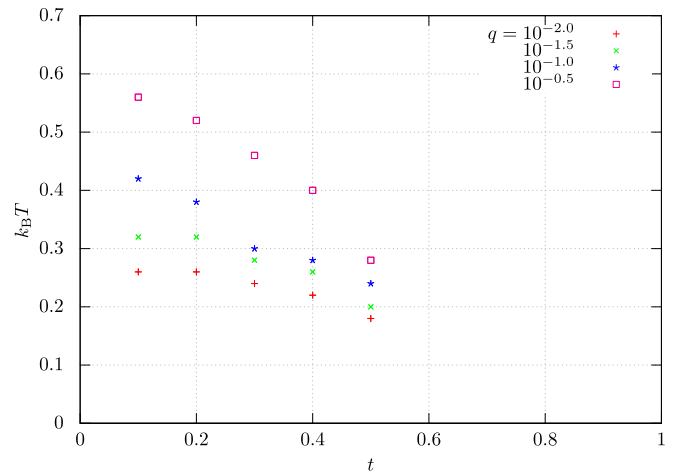


FIG. 48. Phase diagram for $K = 20$, $J = 1.0$, and $\epsilon = 0.00$. The horizontal and vertical axes represent the parameter t , which controls the probability of selecting production rules, and the critical temperature $k_B T$, respectively.

b. $k_B T$ - t phase diagram

The $k_B T$ - t phase diagram is shown in Fig. 48. In the $k_B T$ - t phase diagram, we observe a behavior where the transition

temperatures for different values of q show a tendency to converge toward around 0.2 as t approaches 0.5.

-
- [1] K. Huang, *Statistical Mechanics*, 2nd ed. (Wiley, New York, 1987).
 - [2] N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (Addison-Wesley, Reading, MA, 1992).
 - [3] L. D. Landau and E. M. Lifshitz, *Statistical Physics, Part I*, 3rd ed. (Pergamon Press, Oxford, 1980).
 - [4] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, Scaling laws for neural language models, [arXiv:2001.08361](#).
 - [5] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, *et al.*, Emergent abilities of large language models, [arXiv:2206.07682](#).
 - [6] E. De Giuli, Random language model, *Phys. Rev. Lett.* **122**, 128301 (2019).
 - [7] K. Nakaishi and K. Hukushima, Absence of phase transition in random language model, *Phys. Rev. Res.* **4**, 023156 (2022).
 - [8] F. Lalegani and E. De Giuli, Robustness of the random language model, *Phys. Rev. E* **109**, 054313 (2024).
 - [9] Y. Toji, J. Takahashi, V. Roychowdhury, and H. Miyahara, Berezinskii-Kosterlitz-Thouless transition in a context-sensitive random language model, *Phys. Rev. E* **113**, 015305 (2026).
 - [10] V. Berezinskii, Destruction of long-range order in one-dimensional and two-dimensional systems having a continuous symmetry group. I. Classical systems, *Sov. Phys. JETP* **32**, 493 (1971).
 - [11] V. Berezinskii, Destruction of long-range order in one-dimensional and two-dimensional systems possessing a continuous symmetry group. II. Quantum systems, *Sov. Phys. JETP* **34**, 610 (1972).
 - [12] J. M. Kosterlitz and D. J. Thouless, Ordering, metastability and phase transitions in two-dimensional systems, *J. Phys. C* **6**, 1181 (1973).
 - [13] J. M. Kosterlitz, Kosterlitz-Thouless physics: A review of key issues, *Rep. Prog. Phys.* **79**, 026001 (2016).
 - [14] J. M. Kosterlitz, The critical properties of the two-dimensional model, *J. Phys. C* **7**, 1046 (1974).
 - [15] F. J. Dyson, Existence of a phase-transition in a one-dimensional Ising ferromagnet, *Commun. math. Phys.* **12**, 91 (1969).
 - [16] F. J. Dyson, Non-existence of spontaneous magnetization in a one-dimensional Ising ferromagnet, *Matematika* **16**, 113 (1972).
 - [17] F. J. Dyson, An Ising ferromagnet with discontinuous long-range order, *Commun. Math. Phys.* **21**, 269 (1971).
 - [18] Y. Tomita, Monte Carlo study of one-dimensional Ising models with long-range interactions, *J. Phys. Soc. Jpn.* **78**, 014002 (2009).
 - [19] J. Martínez-Herrera, O. Rodríguez-López, and M. Solís, Critical temperature of one-dimensional Ising model with long-range interaction revisited, *Physica A* **596**, 127136 (2022).
 - [20] F. Y. Wu, The Potts model, *Rev. Mod. Phys.* **54**, 235 (1982).
 - [21] M. A. Harrison, *Introduction to Formal Language Theory* (Addison-Wesley, Reading, MA, 1978).
 - [22] G. Rozenberg and A. Salomaa, *Handbook of Formal Languages, Vol. 3: Beyond Words* (Springer Science & Business Media, New York, 2012).
 - [23] P. Linz and S. H. Rodger, *An Introduction to Formal Languages and Automata* (Jones & Bartlett Learning, Burlington, MA, 2022).
 - [24] A. Meduna, *Formal Languages and Computation* (Auerbach Publishers, Incorporated, Boca Raton, FL, 2014).
 - [25] D. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2021).
 - [26] J. Binney, N. Dowrick, A. Fisher, and M. Newman, *The Theory of Critical Phenomena: An Introduction to the Renormalization Group* (Clarendon Press, Oxford, England, 1992).
 - [27] M. Newman and G. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, England, 1999).
 - [28] G. Jäger and J. Rogers, Formal language theory: Refining the Chomsky hierarchy, *Philos. Trans. R. Soc. B* **367**, 1956 (2012).
 - [29] L. M. Tuan, T. T. Long, D. X. Nui, P. T. Minh, N. D. Trung Kien, and D. X. Viet, Binder ratio in the two-dimensional q -state clock model, *Phys. Rev. E* **106**, 034138 (2022).
 - [30] M. Hasenbusch, The Binder cumulant at the Kosterlitz-Thouless transition, *J. Stat. Mech.* (2008) P08003.
 - [31] H. W. Lin and M. Tegmark, Critical behavior in physics and probabilistic formal languages, *Entropy* **19**, 299 (2017).
 - [32] V. Ardourel and S. Bangu, Finite-size scaling theory: Quantitative and qualitative approaches to critical phenomena, *Stud. Hist. Philos. Sci.* **100**, 99 (2023).
 - [33] Y.-D. Hsieh, Y.-J. Kao, and A. W. Sandvik, Finite-size scaling method for the Berezinskii-Kosterlitz-Thouless transition, *J. Stat. Mech.* (2013) P09001.
 - [34] Z. Zuo, S. Yin, X. Cao, and F. Zhong, Scaling theory of the Kosterlitz-Thouless phase transition, *Phys. Rev. B* **104**, 214108 (2021).
 - [35] H. Ueda, K. Okunishi, K. Harada, R. Krčmár, A. Gendiar, S. Yunoki, and T. Nishino, Finite- m scaling analysis of Berezinskii-Kosterlitz-Thouless phase transitions and entanglement spectrum for the six-state clock model, *Phys. Rev. E* **101**, 062111 (2020).