

# Syntactic Parameters and Spin Glass Models of Language Change

KARTHIK SIVA, JIM TAO, MATILDE MARCOLLI  
*California Institute of Technology*

## Abstract

We develop a model of language change, based on syntactic parameters treated like spin variables in a spin glass model, where the vertices of the underlying graph represent a set of languages, and edges connecting them represent language interaction. As a measure of the strength of language interaction we use data from the MIT MediaLab, while for syntactic parameters we use the SSWL database and Table A of [22]. We develop a new type of spin glass model, based on Ising and Potts models on a graph coupled at the vertices, where the vertex interactions encode entailment relations between syntactic parameters. We study the dynamics of the system via Monte Carlo simulation and a version of the Metropolis–Hastings algorithm. We find that, for independent parameters with no entailment relations, in the low temperature range the parameters tend to align over time according to the most prevalent value in the initial configuration. While in the case with entailment more interesting equilibrium configurations appear depending on the strength of the entailment relation: we focus in particular on the cases of the Strong Deixis and Strong Anaphoricity and the Partial Definiteness and Definiteness Checking parameters. We interpret the temperature variable of the spin glass model as a measure of whether syntactic parameters are best considered as deterministic or probabilistic variables.

## 1. Introduction

How languages change through the interaction of their speakers is a topic of interest to linguists but, like many interacting many-body problems, it is difficult to study it analytically. In this article we follow an approach that views languages at the level of syntax (or morpho-lexicon), with syntactic structures encoded as a vector of binary syntactic parameters, a point of view originating in the Principles

and Parameters model of Generative Linguistics, [5, 6], (see also [1] for a more expository account). In contemporary Minimalism all parameters are postulated to be functional morpho-lexical, rather than syntactic as in the original Principles and Parameters (Government and Binding) theory. Also the traditional distinction between deep and surface structure is considered superseded within the Minimalism framework, but for expository reasons we will adopt the more traditional Principles and Parameters terminology. Our approach and our mathematical analysis are independent of the “location” of parameters. However, the data we use for our simulations reflect the classification of parameters as presented in specific sources: the SSWL database [40] for the independent parameters case, and the data collected in Table A of [22] for the entailed parameters case. In this article we will simply adopt the traditional terminology and loosely refer to “syntactic parameters” for all of these data. We refer the reader to the other contributions in this volume for a more extensive discussion of the nature of parameters, especially [21] and [32].

The data of [22] are generally preferable to the SSWL data, both at the level of the choice of parameters, which more closely represent true syntactic parameters than the binary variables considered in the SSWL data, and also because they record explicit dependencies (which we refer to here as “entailment relations”) between parameters. On the other hand, the SSWL data are at present more extensive, although the data of Table A of [22] have more recently been refined and greatly expanded, see [21].

It is known that syntactic parameters can change in the course of the historical development of a language. Instances of parameter flipping have been identified in the case of some Indo-European languages, see for example [38]. For recent results on language change from the point of view of syntactic parameters, see [15].

The present study is intended to be a more linguist-friendly and mathematically light version of the preprint [35] by the same authors. We show how to construct a model of language change based on three main inputs:

- data of syntactic parameters from the SSWL and Terraling databases [40] and [41],
- a measure of language interaction provided by estimates of bilingualism obtained from MIT MediaLab data [30],
- Statistical Physics methods based on the theory of spin glass models.

We discuss the results of computer simulations carried out in simple versions of this model, in the case of independent syntactic parameters, and in the case of parameters that are related by an entailment relation.

The physical model we use has two additional variables: a *temperature variable*  $T$  and a *coupling constant*  $E$ , which we refer to as “entailment energy,” which is absent (equal to zero) in the case of independent syntactic parameters. The role of the temperature is discussed at length in §1.2.3. In essence, varying the temperature  $T$  in the model allows for the possibility that the expression of parameters is only given statistically by a frequency  $P$  rather than being set uniquely to either a +1 or a -1 value. The zero-temperature case corresponds to the probability  $P$  being zero or one (that is, a uniquely set value for each parameter), while a high temperature corresponds to  $P$  approaching 1/2 (that is, the parameters take either value with equal frequency). In a similar way, the “entailment energy,” variable  $E$ , describes how strongly enforced a given entailment relation between syntactic parameters is in the model. When  $E$  is very large the relation is strongly enforced, while when  $E$  is small it is only weakly enforced, and there is no entailment when  $E$  is zero.

What we refer to here as “entailment” may be more generally referred to as “inference” or “constraint.” We focus on one particular form of entailment between two parameters, which we describe more explicitly in §2.5 below. More general entailment relations involving several parameters can also be accommodated by a modification of the model we describe in this article, through the introduction of suitable, more complicated, interaction terms at vertices in the spin glass model. Our main linguistic application in the entailed parameter case will be a simulation of the dynamics (on a small graph of languages) of the Strong Deixis and Strong Anaphoricity parameters and the Partial Definiteness and Definiteness Checking parameters.

In the rest of this introductory section we give a brief non-technical introduction to spin glass models and we clarify our linguistic assumptions. In the next section we review the main results of [35], again in a non-technical form. We present the actual mathematical model only in Appendices A and B, which can be skipped, without affecting the comprehension of the rest of the paper, by readers who are primarily interested in the linguistic perspective.

## 1.1 What are spin glass models?

Spin glass models are a very versatile set of dynamical models widely used in Statistical Physics to simulate evolutions of systems exhibiting phase transitions and a range of behaviors from chaos to ordered phases, depending on a variable temperature parameter. Spin glass models have also been widely used in the mathematical modeling of neural networks. A simple nontechnical introduction to the theory of spin glasses can be found in [37], while a standard more detailed reference is, for instance, [14]. A mathematical analysis of the main types of spin glasses that we will use as building blocks for the system we consider in this investigation can be found in [13] and [36].

The typical setting for a spin glass model consists of a network (a graph) where at each node one has a spin variable, while edges connecting nodes carry the interaction between the spins. A spin is a variable that can take only a finite number of values. Usually it is assumed to be a binary variable, like a switch, that can only take two values, either up or down. This case is also known as the Ising model. However, it is often useful to also consider cases where the spin variables can take more than two positions. Such multiplicity (usually limited to three possible values) has also been suggested in Linguistics for some parameters.

These generalizations of the Ising model are called Potts models with  $q$  spin states. We will see that, for the purpose of modeling syntactic parameters, it will be best to work with a combination of Ising models ( $q = 2$ ) and Potts models with three possible spin states ( $q = 3$ ). In a typical spin glass model, the interaction between spin variables located at nearby vertices of the graph tends to push the two spins toward alignment, in the sense that the configuration of aligned spins is favored (lower energy) over the one with anti-alignment. This is called the ferromagnetic case, since the typical physical example of a system of spins that tend to align arises in ferromagnetic materials — these are magnetized at low temperatures, while at higher temperatures they lose their magnetization. The anti-ferromagnetic case has the opposite behavior: the interaction between nearby spins tends to turn them into an anti-aligned position. In our setting, we will work with a model with only ferromagnetic interactions, although we will discuss briefly the question of a possible role for the anti-ferromagnetic ones. The behavior of a spin glass

system subject to ferromagnetic interactions depends on a temperature parameter. At low temperatures, typically the system settles on one of two possible equilibrium states with either all the spins in the down positions, or all the spins in the up positions, while at higher temperatures one sees a disordered phase with randomly distributed spins and no non-trivial overall average magnetization. This is the typical behavior exhibited in nature by ferromagnetic materials. There are efficient algorithms to simulate the dynamics of Ising and Potts models, in particular Monte Carlo methods and the Metropolis–Hastings algorithm. The latter is a discretized gradient descent method aimed at flowing towards the minima of the energy functional. This algorithm works by generating a Markov chain process, for which an ergodicity result (namely, the property that the dynamics is sufficiently mixing) is required to ensure the stability of the process and the convergence of the dynamics to a unique stationary distribution. Indeed, the ergodicity property guarantees that the dynamics of the model can reach all possible configurations. For the Ising case, all these facts are described in great detail in [18] and [27]. In the case of Potts models with more than two spin states ( $q > 2$ ), a version of this method and the necessary mathematical properties are discussed in [2] and [11]. We will use a version of the Metropolis–Hastings algorithm to obtain computational simulations of the dynamics of our model of language change.

What is new in our approach, from the point of view of spin glass models, are the following two results described in Appendix A:

- We construct a new type of spin glass model, different from those usually considered in the physics literature, given by a family of Ising and Potts models on the same underlying graph, coupled by interaction terms at the vertices. We use it to describe the dynamics of entailed syntactic parameters. This type of spin glass model reduces to the known case of a Potts model with external magnetic field when all but one of the parameters are frozen (rendered non-dynamical).
- We extend the usual proof of ergodicity that allows for the use of the Metropolis–Hastings algorithm to simulate the dynamics of the spin glass model, originally proved for Ising models on a lattice, to the case of the models we introduce here. This allows us to obtain reliable estimates on the time evolution of the system using a version of Metropolis–Hastings.

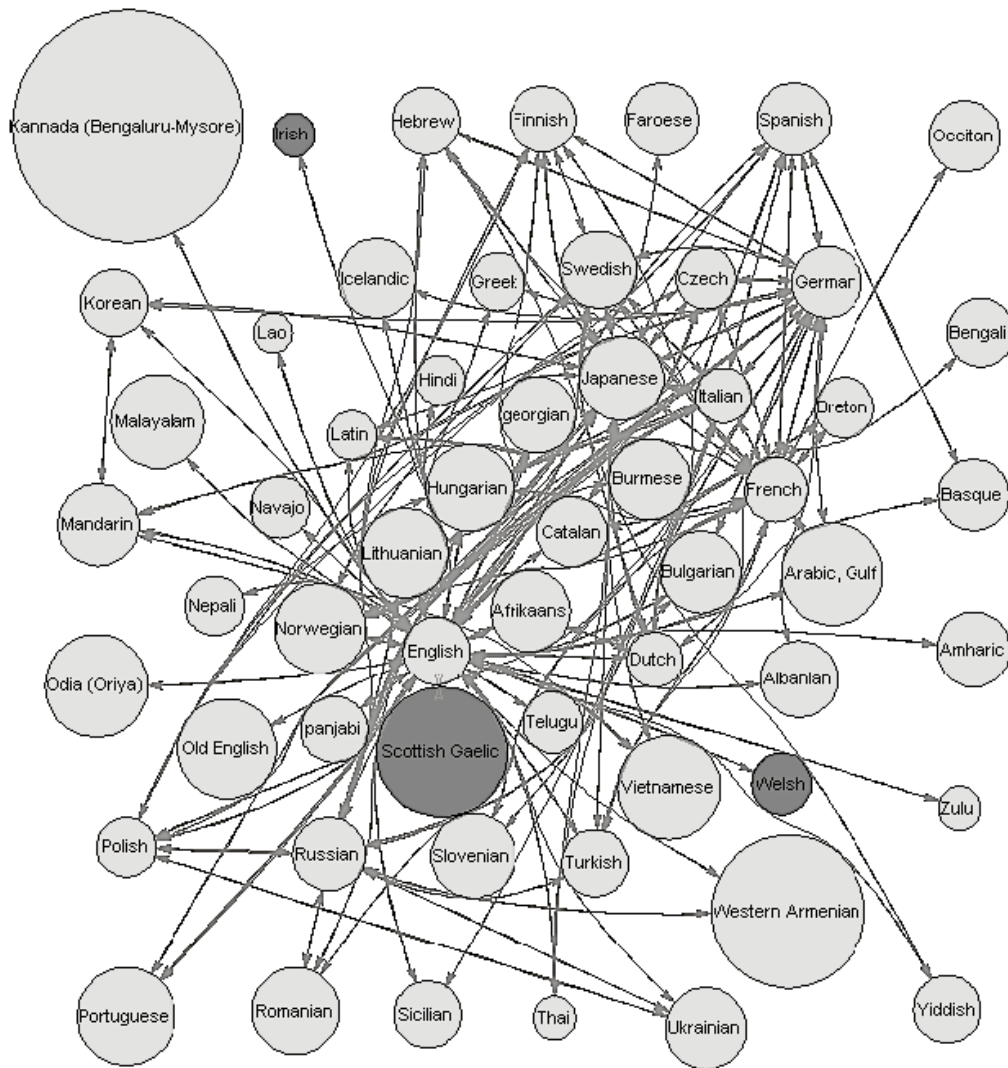


Figure 1. Initial state of the Subject-Verb parameter for various languages in the SSWL database. Most languages currently possess this syntax (value +1 of the parameter, red colored vertices). Graph built on MIT Media Lab data of language interaction. (Vertex sizes are for visualization purposes only.)

Note that, in itself, the fact of considering several coupled spin glass models on the same underlying graphs is not new; it is used, for instance, in the framework of Markov Random Fields [33]. However, the specific type of interaction terms, which we use to describe the entailment relations between syntactic parameters, is new to this work.

## 1.2 Linguistic assumptions

We make a series of simplifying linguistic assumptions, for the purpose of obtaining a computationally feasible model. We examine here the plausibility of these assumptions and their interpretation from a Linguistics point of view.

### *1.2.1 Languages and parameters*

First, we assume that the languages we simulate are sufficiently distinct and never merge. We do not concern ourselves with whether a dialect of a language is truly distinct from that language, or whether two closely related languages may at some point merge into the “same” language. Instead, our model assumes that there exist reliable descriptive criteria to make a relevant distinction between languages. Since we use data from the SSWL database for language parameters and from the MediaLab database [30] for language interaction, we take the world languages that are listed in both databases.

We only analyze languages at the level of parametric differences. In particular, we do not keep track of any information on mutual influence on single words in the non-functional component of the lexicon.

Considering languages as “discrete” objects (as opposed to a continuum of dialects) is a rather common linguistic assumption. Alternative models, such as wave models of transmission of linguistic changes are also possible (see for example [26]), but we will not consider them in this paper. It would be interesting to see whether statistical physics methods could be relevant to wave models of languages, but that is outside the purpose of our present investigation.

### *1.2.2 Independent parameters and entailed parameters*

A second simplification we make is that, for a given syntactic parameter, such as the “Subject-Verb” and other word order parameters listed in SSWL, a language either has it expressed or not. However, even though we assume parameters to be in a binary state (either zero or one) and not in a mixture of the two, we will accommodate the possibility of noise and of a frequency  $P$  of expression that can be between zero and one, through the temperature variable of the spin glass model.

Indeed, one can account for finer syntactical structures where the values of the parameters are only interpreted statistically, as frequencies (or probabilities), as in [20], for instance. This approach can be used to partially correct for confluences of deep and surface structure in some databases of syntactic parameters (this will be discussed in more detail in §2.3 and §1.2.3 below, where we show how one can accommodate this point of view in our model through the role of the temperature variable of spin glass dynamics).

We also allow for the possibility that a parameter may simply be undefined for a given language due to a pairing relation (which we refer to as entailment) between different parameters.

For example, an entailment relation happens when there are pairs of parameters  $(p, p')$  with the property that if  $p = +1$  then  $p'$  is undefined, while if  $p = -1$ , then  $p'$  can take either value  $\pm 1$ . Several specific examples of such entailment relations are identified in [22–23]. The more extensive new data presented in [21] also record detailed information on entailment relations between parameters.

We distinguish between two different cases, for a dynamical model involving several syntactic parameters:

- *Independent parameters*: This is the case of parameters that behave like independent binary variables. This means that the dynamical evolution of each given parameter does not depend in any way on the values of the other parameters. The dynamics depends only on the initial value for that given parameter (for all the various languages on the graph) and on the strengths of the interactions along the edges of the graph.
- *Entailed parameters*: This is the case where there are relations between parameters. These parameters now behave as coupled variables in the dynamical evolution of the system, and their respective values and relations also affect the dynamics.

In particular, in our model, the presence of entailment relations alters the independent parameters dynamics in two ways:

- an additional possible value 0 of the parameters should be introduced, which accounts for the fact that at least one of the related parameters can become undefined by entailment from other parameters.

- interactions between parameters should be introduced in order to model the entailment property.

We describe in more detail in §2.5 below how this is achieved by modifying the classical Potts models by introducing a new system given by a family of Potts models on the same graph, coupled at the vertices.

### *1.2.3 Parameters as deterministic or statistical variables*

In Linguistics, a probabilistic approach to syntactic parameters was recently developed in [20]. This work makes a strong case, based on an analysis of treebanks of different sizes, for a sample of 20 languages, for viewing the syntactic parameters of a given language not as frozen in the up or down position, with binary 0 or 1 values, but as a binary probability distribution  $\{P, 1 - P\}$ , for some  $0 \leq P \leq 1$ , which expresses the *tendency* of the language to have a given syntactic parameter expressed or not:  $P = 1$  if it is always expressed,  $P = 0$  if it is never expressed, and an intermediate value of  $P$  if it is expressed with an overall frequency  $P$ . For example, the case of the Head-Initial structure is analyzed in [20] and it is shown that languages typically tend to express elements of both 0 and 1 possibilities for this parameters, with a certain probability. Similar results are obtained, in the same article, for Subject-Verb, Object-Verb, and Adjective-Noun parameters.

In our spin glass model, a similar role is played by the temperature variable of the spin glass. In a physical system, while at zero temperature  $T = 0$  all the spins are frozen in either the up or down position; at higher temperatures they fluctuate between the up and down state, with a certain frequency (which depends on the language and the parameter). Making all these frequencies depend on a single temperature variable is clearly an approximation we are making that simplifies the model. We discuss in more detail in Appendix A how the frequencies (in the sense of [20]) with which parameters are expressed in the various languages can be modeled as functions of the temperature variable of the spin glass.

Note that several different interpretations of the role of the physical temperature parameters can be proposed for a spin glass model of language change. Some interpretations, in a scenario of interaction between a small set of languages, would relate temperature to

historical and social phenomena that accompany the interaction of populations. However, such phenomena can more appropriately be accounted for in adjusting the interaction energies along the edges of the graph, which represent the strength of the interaction (ferromagnetic or antiferromagnetic) between a given pair of languages. The overall temperature variable, on the other hand, allows for more or less noise in the determination of the parameter values.

Indeed, as we discuss in more detail in §2.3 below and in Appendix A, we propose here to think of the temperature parameter as a way to account for noise in the parameter data, either due to the existence of a possible probabilistic structure, or as a way to compensate for uncertainties created in parameter databases by the conflation of deep and surface structure. While the distinction between deep and surface structure is no longer considered within the Minimalism model, in our exposition here we refer to the traditional terminology of the Principles and Parameters model.

This approach to interpreting the role of the temperature parameter and introducing noise in the determination of syntactic parameters is also consistent with our model of language interaction based on bilingualism. Typically, bilingual populations present very interesting Code-switching behavior, see for instance the analysis in [24] and [39]. The level of noisiness in the determination of parameters, defined in terms of frequencies  $P$  with  $0 < P < 1$ , regulated by the overall temperature parameters (as discussed in Appendix A), then can be seen as measuring the frequency with which, in a bilingual population, Code-switching introduces syntactic constructions from one language in the other, carrying along the corresponding configuration of parameters.

#### *1.2.4 Language interaction*

Our model aims at investigating how syntactic parameters may evolve in time, assuming that interaction between different languages tends to favor configurations with aligned values of the parameters, as in ferromagnetic spin glasses.

A way of interpreting the geometry of our model is to imagine our graph of languages as depicting various populations (the vertices of the graph) with a record of the syntactic parameters of the main language they speak, and with edges between the vertices carrying a record of the amount of linguistic interaction between

these populations, measured in terms of numbers of translations as in the MediaLab study [30]. These data on the mutual influence of languages are based on an estimate of the amount of bilingualism, measured in different ways such as number of book translations, or on the likelihood that a Wikipedia editor who edits in one language would also edit in another. Clearly, languages that are more likely to influence others tend to be those that are dominant for political and economic reasons, more diffuse and commercially useful, and these facts are reflected in the data.

For example, Figure 1 depicts the initial state of our language graph for one of the parameters considered—the Subject-Verb parameter. The vertices are represented as circular nodes colored light gray if the language possesses the parameter, or dark gray if it lacks the parameter. The sizes of the vertices do not signify anything in these graphs: variable sizes are only used for ease of visualization.

Since our model of interaction energies for the spin glass system is based on measures of bilingualism, as obtained by Media Lab, the dynamical change of syntactic parameters that we describe would be best understood linguistically within the context of the theory of bilingual Code-switching, [24]. For example, as observed in the study [39] of an English–Spanish bilingual population, the language spoken by this population “acquires” the Pro-Drop parameter (which is +1 in Spanish but -1 in English) in the sense that Spanish verbs inserted in English sentences retain their Pro-Drop form from Spanish. In the dynamical change of syntactic parameters we obtain from the spin glass simulation, we see previously inactive -1 parameters switch to the activated form +1 as an effect of “ferromagnetic” alignment with the corresponding parameters of nearest neighbor vertices. In a realistic model, this should be interpreted (at least in a first phase of language change) in the sense of Code-switching theory, as imported forms from the neighboring languages that are inserted in the original syntactic structure, while retaining their parameters activated.

Although we do not discuss them explicitly in this article, it is possible to construct, with the same method described here, models where some of the language interactions are “anti-ferromagnetic” rather than “ferromagnetic.” One can imagine situations where interaction between the languages spoken by two populations may favor maintaining or increasing the difference between the two languages, for instance when political tension and hostile relations make cultural and linguistic differentiation more desirable.

### 1.2.5 Language change versus evolution

To clarify our use of terminology: the model we propose in this article concerns *language change* over time, seen at the level of syntactic parameters, as a result of the interaction between different languages which is achieved through the presence of bilingual populations. Models of language change have often been proposed in terms of *language evolution*, see for instance [7], [10], [28]. This evolutionary approach has often been criticized as inappropriate for the modeling of historical language change on the basis of the fact that it suggests mechanisms analogous to biological evolution, including but not limited to inheritance, reproduction and selection, which do not necessarily apply to natural languages, see for instance [12] and [17]. In this article we will occasionally use the term *evolution* in a completely different sense, which simply refers to the dynamics, or “time evolution” of a physical system, completely unrelated to any biological meaning.

## 1.3 Quick summary of the main results

We ran computer simulations of language change based on a spin glass model with an initial configuration of syntactic parameters taken from the SSWL database [40] and from Table A of [22], and with language interaction strengths taken from bilingualism data from the MIT MediaLab study [30]. The spin glass model is an Ising model for independent syntactic parameters and a coupling at vertices of Ising and Potts models for the case of syntactic parameters paired by entailment relations.

The dynamics depends on a physical temperature parameter, whose linguistic significance we discuss in §2.3 and §1.2.3 and in Appendix A.

In the low temperature regime, and in the case of independent parameters, we find that the prevalence of the expressed +1 or non-expressed -1 state in the initial configuration for the given set of languages determines the final equilibrium configuration with the parameter aligned in all languages to the most prevalent initial state. The system converges very rapidly to the equilibrium state. On the other hand, in the high temperature regime, the state of the given parameter continues to oscillate during the dynamics as a result of the interactions, and in the final equilibrium the system fluctuates around configurations with on average half of the languages with

the parameter in +1 and half in the -1 state. This behavior can be read off Figure 4, where it is plotted for the dynamics of the single Subject-Verb parameter, with initial condition taken from the SSWL database. The first plot of Figure 4 shows the simulated dynamics for the low temperature regime; one can see that within very few steps of the algorithm the system is already stable and in the final equilibrium state, with the parameter in the +1 position (the most prevalent in the initial configuration, as shown in Figure 1). This means that the final state, in the low temperature regime, looks like Figure 2 and is reached very rapidly. The second plot of Figure 4 shows the simulated dynamics for the high temperature regime; in this case, even after a large number of steps in the algorithm, the configuration of the system is widely oscillating, which means that the values of the parameter for the various languages will keep changing. These oscillations are centered at the value zero, which means that on average the system is in a state where half the parameter values are +1 and half are -1; that is, the equilibrium configuration looks like Figure 3.

In the case of pairings of entailed parameters, we focus on the case of the pair of Partial Definiteness and Definiteness Checking parameters and on the pair of Strong Deixis and Strong Anaphoricity parameters, among those listed in Table A of [22], and we run simulations on very small graphs of languages for which we can prove the algorithm has the desired mathematical properties.

In the case of such pairings of parameters, the dynamics of the system depend to two variables: the temperature parameters  $T$ , as before, and an additional variable  $E$  that expresses how strongly the entailment relation that pairs the two parameters is enforced. When we run the simulation of the dynamics in this case, we find four different scenarios corresponding to the regimes with large  $T$  and large  $E$  (HT/HE), small  $T$  and large  $E$  (LT/HE), large  $T$  and small  $E$  (HT/LE), and small  $T$  and small  $E$  (LT/LE).

For the pair of parameters Partial Definiteness and Definiteness Checking, with the initial configuration taken from the data of Table A of [22], one can read the results of the dynamics in Figure 5 for the HT/HE regime (top row) and for the LT/HE regime (bottom row) and in Figure 6 for the HT/LE regime (top row) and for the LT/LE regime (bottom row).

Comparing the top rows of Figure 5 and Figure 6, one sees that, when the temperature is sufficiently high, the dynamics continues to exhibit a sequence of random oscillations, even after a very large

number of steps in the algorithm. This random oscillatory behavior is similar to what we found in the high temperature case of a single independent parameter in the second plot of Figure 4. However, we see now that the pattern of oscillations is different, depending on whether the other variable  $E$  is large or small.

When the entailment relation is only weakly enforced, the values of both Partial Definiteness and Definiteness Checking fluctuate around an equilibrium where both parameters are defined (values  $\pm 1$ ), while the entailment relation (which makes the second parameter undefined when the first has  $-1$  value) is completely erased by the dynamics. One can see this type of equilibrium in the tables in §2.7.1, in an example based on a very small graph of three interacting languages, by comparing the initial configuration for the Partial Definiteness and Definiteness Checking parameters in the first table, with the equilibrium configuration in the second column of the third table. This situation can occur, for example, in a bilingual population, when the entailment relation is enforced in one language but not in the other, hence it occurs only with a certain frequency in Code-switching, when syntactic construction from one language are inserted in the other (see for instance the analysis of Code-switching in [24], [39]).

When the entailment relation is strongly enforced, in the final equilibrium configuration around which the dynamics oscillates, the entailed parameter tends to acquire the undefined value  $0$  even when the other parameter settles on the  $+1$  value that would allow the entailed parameter to be defined with either  $\pm 1$  values. This can be seen by comparing the first table in §2.7.1 (initial condition for the parameters) with the first column of the third table.

In the low temperature regime (second row of Figure 5 and Figure 6) one sees that the dynamics very quickly converges to equilibrium. As in the case of a single parameter, the equilibrium configuration has all languages align their parameters to a common value. However, which common value is selected as the equilibrium configuration now depends on the strength of the entailment relation. In the large  $E$  case, the equilibrium configuration preferentially selects the value where the entailed parameter is undefined (third column of the third table in §2.7.1), while in the low  $E$  case the equilibrium configuration has both parameters defined and in alignment with the prevalent  $+1$  position of the entailing parameter in the initial configuration.

For the pair of parameters Strong Deixis and Strong Anaphoric-

ity, also with initial configuration from the data of Table A of [22], the results of the dynamics simulation are given in Figure 7 for the HT/HE regime (top row) and the LT/HE regime (bottom row) and in Figure 8 for the HT/LE regime (top row) and the LT/LE regime (bottom row).

In this case, again, we see that in the high temperature regime (top row Figure 7 and Figure 8) random oscillations persist in the dynamics. When analyzed on a small graph of four interacting languages, we again find that the equilibrium configuration in the large  $E$  case tends to increase the number of languages where the entailed parameter becomes undefined, even if the entailing one remains in the +1 position, while in the low  $E$  case the equilibrium configuration erases the entailment relation and leaves both parameters in a defined (either  $\pm 1$ ) position.

In the low temperature case (bottom row Figure 7 Figure 8), the dynamics very rapidly converges to an equilibrium position. However, in this case the equilibrium does not necessarily consist of a configuration where all languages have aligned parameters. Indeed, we find a case (fourth column of the fourth table in §2.7.1) with an equilibrium where half of the languages have acquired the entailment relation with the entailed parameter undefined, while the others have both parameters defined and aligned to (+1,-1) values.

These examples are not very realistic, because they only involve a choice of a very small number of languages and they ignore the significant presence of interaction with other languages outside of the ones chosen. However, they suffice to show that different more complicated equilibrium configurations are possible when entailment relations between parameters are taken into consideration. A more detailed study over a larger graph of languages and with more complex combinations of interrelated parameters is ongoing.



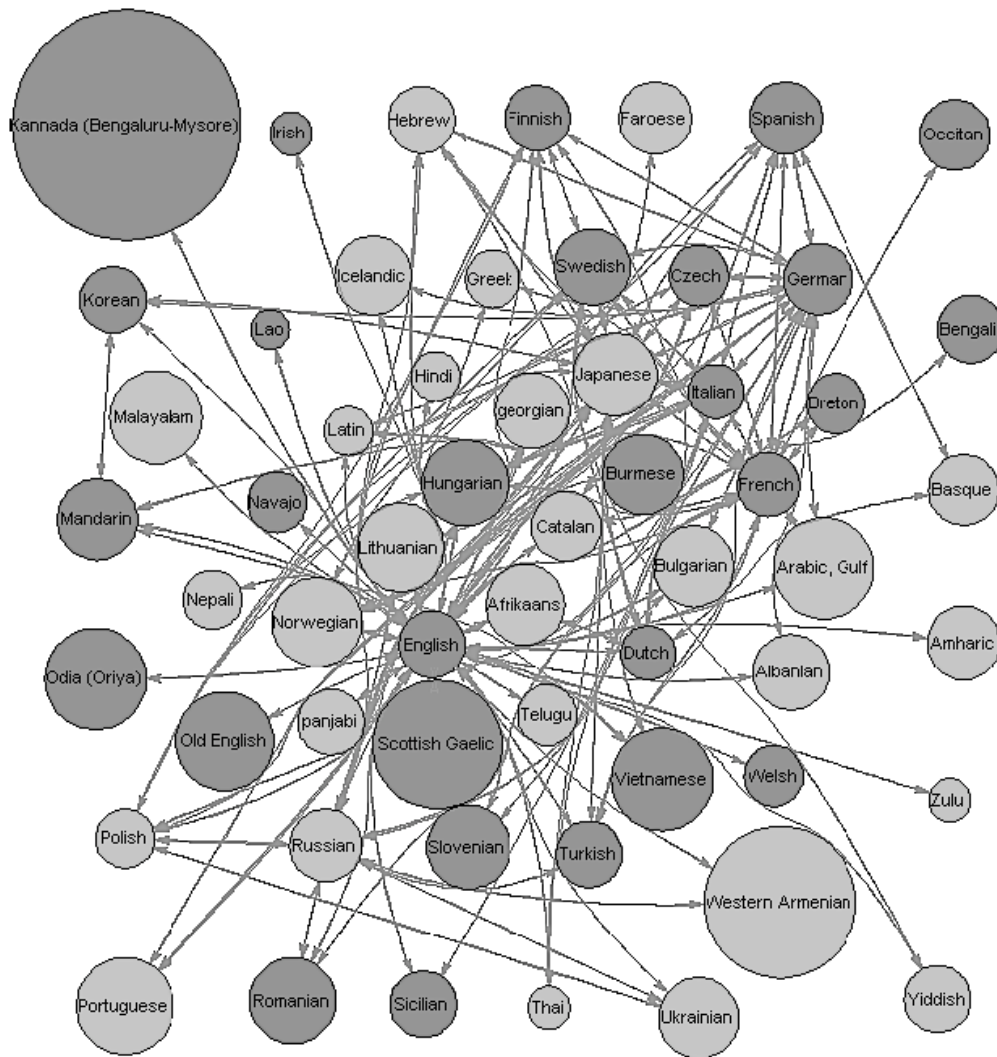


Figure 3. In the high temperature regime equilibrium for the Subject-Verb parameter (here  $T = 20$ ), all vertices have local magnetization close to zero, with half of them approaching zero from the positive direction.

## 2. The main construction

### 2.1 Syntactic parameters as spin variables

We consider a list of 58 languages belonging to a range of different language families. These are the languages that are listed both in the SSWL database of syntactic parameters of world languages [40] and in the MIT MediaLab database [30]. They are the languages listed at the vertices of the graph of Figure 1. To each language we assign a

binary list of syntactic parameters, as recorded in the SSWL database. We consider each binary syntactic parameter as a spin variable, with two possible values  $\pm 1$  — with value  $+1$  if the parameter is expressed in the language and  $-1$  if it is not. Thus, for each language we have a collection of spin variables — one for each syntactic parameter — for which we take as initial condition the values of these parameters as recorded in the SSWL database.

Entailment relations between parameters provide more subtle information, as these relations are not recorded explicitly in the SSWL database. For the purpose of the examples of dynamics with entailment presented in this article, we selected a small group of languages and pairs of entailed parameters taken from the lists of parameters reported in [22] and [23]. In the case of entailed parameters, we not only consider binary values  $\pm 1$  for the parameters, but we also consider that those parameters that can be made undefined by entailment from other parameters should have three possible states  $\pm 1$  and  $0$ , where the value  $0$  means that the parameter is undefined. Thus, in the presence of entailment relations, we assign to a language a list of spin variables, one for each syntactic parameter, where some of the spin variables have two possible values  $\pm 1$ , while some other variables have three possible spin states,  $\pm 1$  and  $0$ .

## 2.2 The graph of languages

We identify each language with a vertex of a directed graph, and each interaction of the form “language A influences language B” with a directed edge from vertex A to vertex B. Such an edge has a weight or interaction strength  $J_{AB}$  given by the aforementioned metric used by MIT Media Lab. Note that in general,  $J_{AB} \neq J_{BA}$ . The language interaction can be symmetrized by considering the total amount of mutual interaction given by the sum  $J_{AB} + J_{BA}$ .

As explained above, we represent the presence or absence of a particular syntactic parameter in a language by associating a spin variable with the corresponding vertex, where a spin of  $+1$ , “spin-up,” indicates the former and  $-1$ , “spin-down,” represents the latter. The physical analog of this arrangement is the association of a particle with spin attributes to each vertex and inter-particle interactions with the edges, the usual setting of the Ising model in statistical physics. The resulting graph of languages, showing the initial configuration of the Subject-Verb syntactic parameter, is shown in Figure 1.

### 2.3 Interpretation of physical parameters

Our physical model depends on two variables: a temperature parameter  $T$  and an interaction parameter  $E$ . The temperature parameter has the usual role as in the Ising and Potts models, while the parameter  $E$  arises as strength of the coupling of different Ising and Potts models at the vertices of the same graph. It is introduced as a measure of the strength of the entailment relation between parameters. In so doing we adopt a probabilistic approach to syntactic relations, in a way similar to the probabilistic interpretation of syntactic parameters considered in [20]. Namely, instead of just deciding whether a syntactic parameter is entailed from another parameter as a yes/no question, we consider the frequency with which, in a given language, the entailment relation occurs. A high frequency corresponds to a greater strength of the entailment relation (large values of  $E$ ) and a lower frequency corresponds to lower values of the “entailment energy” variable  $E$ .

The temperature variable  $T$  defines the extent to which languages are “noisy” due to external factors (an analog of the heat bath in physics). As discussed in §1.2.3, by “noisy” we mean the fact that the setting of parameters, instead of being a determined value  $\pm 1$ , is interpreted only probabilistically with a probability  $0 \leq P \leq 1$  of the parameter being expressed. A value of  $P$  closer to either 0 or 1 means less noise and a value close to  $1/2$  means greater noise. We think of the temperature variable  $T$  as determining the probability  $P$  (as a function of  $T$ ) so that we can increase or decrease noise in the model by raising or lowering the temperature variable (see Appendix A for a more detailed discussion). If the parameters are individually very noisy, then we can expect that these fluctuations will, on average at equilibrium, result in no overall alignment of a syntactic parameter (magnetization of zero). In this scenario, the languages fluctuate independently. On the other hand, at zero temperature all of the noise is frozen out, so that the system will be found only in the configuration with the largest probability; the energy minimizing ground state, in which all languages either possess or lack a given parameter (magnetization of  $\pm 1$ ).

The possibility of varying the temperature parameter in the model can be used as a possible method to correct for some potential problems inherent in the data of syntactic parameters obtained from databases. For example, in the data presented in the SSWL database,

surface orders are sometime confounded with the deep underlying parameter values. German is verb final with verb-second (V2) effects, while SSWL only codes the surface order. This can create problems with the analysis, see [3]. However, we suggest that this type of problem may be at least in part accounted for by a careful use of the temperature parameter in the spin glass model. For example, in the case of German mentioned above, the verb final parameter will be expressed with a certain frequency  $0 < P < 1$ , where the fact that  $P$  is not equal to one accounts for the correcting V2-effects. This is the type of approach proposed in [20]. A correct value for  $P$  (depending on language and parameter) can be linked to the temperature variable in the way described more in details in Appendix A, so that tuning the temperature parameter achieves a more realistic representation of the behavior of the syntactic parameters. This depends on empirical laws that establish the dependence of the frequencies  $P$  on the temperature  $T$  which require more extensive data to be better understood.

## 2.4 The case of independent syntactic parameters

We first consider the simpler case of syntactic parameters that do not have any entailment relation with other parameters. We call these “independent parameters.” For such parameters, the dynamics is independent of the values of any other parameters; it depends on the initial values of the parameter for the various languages at the vertices of the graph, the interaction energies along the edges, and the temperature variable of the system. The interaction energies along edges are designed to provide a rough measure of the amount of bilingualism, by using data on certain linguistic interactions such as translations; the stronger the interaction between two languages the more the dynamics tends to align the values of a given parameter for the two languages. Thus, for the case of independent parameters, the dynamics can be analyzed a single parameter at a time, modeled by a classical Ising model on the given graph. The only modification required, in order to analyze the system computationally, is to check the validity of the Metropolis–Hastings algorithm for the Monte Carlo simulation, for a graph that is not a lattice. This is discussed in Appendix A.

## 2.5 Syntactic parameters with entailment relations

We now discuss how one can adapt the previous computational setting to the more refined model that also takes into consideration dependencies and entailment of syntactic parameters.

The first modification in the model consists of the fact that we use, for each parameter, a set of three possible values  $\{-1, 0, +1\}$  instead of the binary values  $\{\pm 1\}$ , thus allowing for the possibility that a parameter may be undefined (value 0) by the values of other parameters. Thus, instead of the Ising model, we are looking at a Potts model with  $q = 3$ . The natural generalization of the Metropolis dynamics of the Ising model is the wider class of Glauber dynamics on graphs (see the discussion in Appendix A), where the single-spin-flip dynamics is replaced by the assumption that each move changes the parameter value at just one vertex, by choosing a possible new value uniformly at random. Selection (acceptance/rejection) of the new possible state is then performed on the basis of an energy estimate, where the energy is measured by the Hamiltonian (the energy function) describing the strengths of all the interactions in the system.

In our model the Hamiltonian (that is, the energy function) measures the strength of the mutual interaction between languages and also the strength of the inference (entailment) relation between syntactic parameters. The values of the mutual interaction between languages is an input, obtained from data of bilingualism like those of [30], while the strength of the relations between parameters is a variable in the model which can be tuned to account for weaker or stronger inference, and to reflect how Code-switching in bilingual populations can alter the pairing relations between parameters.

We consider a simple situation, where we have two parameters  $(p_1, p_2)$  with the property that, in a given language, if the first parameter is expressed,  $p_1 = +1$ , then the second parameter  $p_2$  can take either values  $\pm 1$ , while if the first parameter is not expressed,  $p_1 = -1$ , then the second parameter is undefined,  $p_2 = 0$ . As before, we associate spin variables with the two parameters, but now we assume that the first spin variable has only two possible spin states  $\{\pm 1\}$ , as before, while the second spin variable has three possible spin states,  $\{+1, 0, -1\}$ , with 0 for the case when the parameter is undefined.

Thus, we see that, in order to accommodate the entailment relations, we need to construct a multivariable spin glass model, where some of the spin variables associated with the vertices behave like an

Ising model (with number of spin states  $q = 2$ ), while other variables behave like a Potts model (with number of spin states  $q = 3$ ).

Moreover, the entailment relation should be modeled by additional interaction terms in the Hamiltonian, that couple the spin variables of the entailed parameters at all vertices of the language graph, in such a way that the configuration with the first spin variable is equal to  $+1$  and the second equal to  $\pm 1$ , and the configuration with the first variable is equal to  $-1$  and the second being equal to  $0$  are favored energetically over all the other possible configurations.

This leads us into a new class of spin glass models, which to our knowledge has not been considered in physics so far, where different spin variables, with different numbers of spin states coexist on the same graph and are coupled with one another at the vertices, while all are subject to the same interactions along the edges. We describe in more detail in Appendix A how to construct an interaction term that achieves the desired result.

## 2.6 Simulation algorithm

Having established the equilibrium physics expected of a spin glass model of syntax, we will now explore the evolution of the system computationally. Naively exploring the configuration space is computationally too expensive, so instead we will employ a Markov Chain Monte Carlo simulation to propagate stochastically an initial configuration of syntactic parameter values (as specified by the SSWL) toward equilibrium, through the Metropolis algorithm, see for instance [18], [27]. The algorithm and the mathematical assumptions needed for its application are reviewed in Appendix A.

### 2.6.1 Implementation

We obtain the binary values of the syntactic parameters from the SSWL database [40], which documents these values for 115 syntactic parameters and over 200 natural languages. For the case of entailed parameters, we use parameter data from Table A of [22], since the SSWL database does not record any pairing and inference relations between parameters. To model the interaction strengths between languages, we use data from the MIT Media Lab [30], by defining the strength of the influence of language A on language B as the likelihood that two languages are to be co-spoken.

In particular, in their database, they provide two independent measures of interaction: one of them is based on the number of book translations between two languages, while the other considers two languages connected when users that edit an article in one Wikipedia language edition are significantly more likely to also edit an article in another language edition, with the frequency of such editing measuring the strength of the interaction. In one simulation (independent parameters case) we used the Wikipedia edits data as interaction strengths,<sup>1</sup> while in another simulation (entailed parameters) we used the book translation data.<sup>2</sup>

Part of the implementation consisted of parsing and converting the SSWL database of syntactic parameters in a form usable for data analysis. Additionally, discrepancies in nomenclature between the MIT Media Lab topology source and the SSWL database had to be identified and manually resolved. The implementation for the independent parameters simulation was executed in MATLAB. The implementation for the entailment of parameters was executed in Java. The source files are available at the GitHub repository: [https://github.com/pointofnoreturn/spin\\_glass\\_model](https://github.com/pointofnoreturn/spin_glass_model)

## 2.7 Simulation results

Our choice of data here reflects the way parameters are classified in the SSWL database [40], although a similar analysis can be carried out with other initial configurations of parameters. In particular, the SSWL database has several parameters listed that involve surface word order: Subject-Verb, Verb-Subject, Verb-Object, Object-Verb, SVO, SOV, VSO, VOS, OSV, OVS. We refer the reader to [31] for an analysis of the dependencies between these parameters, based on Kanerva networks and recoverability in sparse distributed memories. The results of [31] show different degrees of recoverability for these syntactic parameters in a Kanerva network; this occurs as a superposition of an overall effect which depends on the relative frequencies of expression of the parameters among world languages (which can be seen by using random data with the same frequencies), and an additional effect, independent of frequencies, which should account for actual syntactic dependencies.

<sup>1</sup> <http://language.media.mit.edu/visualizations/wikipedia>

<sup>2</sup> <http://language.media.mit.edu/visualizations/books>

For the purposes of our analysis here, we simplify this situation and consider only the first of this list of word order parameters, ignoring dependencies. How to model the type of dependencies detected in [31] within a spin glass formalism requires further investigation and will be discussed elsewhere.

The results of our simulation show that, in the case of independent parameters, which have no entailment relations with other parameters, the final low temperature equilibrium state of the evolution of the system depends only on what state of the parameters (+1 or -1) is prevalent in the initial state among the languages in the graph. For most syntactic parameters, and for the graph of languages considered, the prevalent state is +1, namely there are more languages that have the parameter expressed than unexpressed. With such initial condition in the low temperature regime, the long term evolution of the system tends to bring all languages into alignment with +1 value of the given syntactic parameter. This is illustrated in Figure 2, in the case of the Subject-Verb parameter, considered without any entailment relation.

By contrast, when the temperature parameter is large, which in our model corresponds to a parameter that is expressed statistically with some higher level of noise, as discussed in §1.2.3 and §2.3, and in Appendix A, we find possible long term equilibria where values  $\pm 1$  of the parameter are distributed over the languages in the graph with zero average magnetization. In this case all vertices have local magnetization close to zero, indicating fluctuations of the value of the parameter, with half of them approaching zero from above. This situation is illustrated in Figure 3. The details of the simulation in the independent parameters case are reported in Appendix A.

There are two points to note here: the model for independent syntactic parameters behaves as predicted for uncoupled Ising models on the same underlying graph. It is marked by the same symmetry-breaking phase transition observed in the 2D Ising Ferromagnet at a critical temperature. This is not surprising, since this model is at its core just a ferromagnet with a more complicated topology.

### *2.7.1 Simulation for entailed parameters*

The dynamics of the system becomes more interesting in the case of entailed parameters. In order to show more clearly the new phenomena that occur in this case, we have run the simulation on a much smaller graph of languages, which exhibit different initial

configurations of the entailed parameters and observed the time evolution of the resulting system.

We consider here a simple case where we focus on two parameters with an entailment relation, over a small set of Indo-European languages. In this case, we take the data on the values of the syntactic parameters from [22], [23], while we still use the same data from the MIT Media Lab database [30] for the interaction strengths between different languages, using the interaction data based on book translations.

We consider, for example two parameters  $p_1, p_2$  expressing Definiteness (respectively, number 7 and 12 in the list in Table A of [22]). The first parameter  $p_1$  expresses Partial Definiteness, while the parameter  $p_2$  is Definiteness Checking. See for instance [9] for a discussion of the role of these parameters. We consider three languages—English, Russian, and Bulgarian. We input interaction energies taken from the same MIT Media Lab data graph, and we neglect the effect of interaction with other languages. The initial configuration of the two parameters, taken from Table A of [22] is according to the following table:

	$p_1$	$p_2$
$l_1$	+1	-1
$l_2$	-1	0
$l_3$	+1	+1

Our second example considers another pair of parameters  $p_1, p_2$ , given, respectively, by Strong Deixis and Strong Anaphoricity (numbers 52 and 53 in Table A of [22]). For a detailed discussion of deixis see [19], while regarding anaphoricity, see for instance [8]. In this example, we work with the complete graph on four vertices (tetrahedron graph), where the vertices  $l_1, \dots, l_4$  correspond to the languages English, Welsh, Russian, and Bulgarian. Notice that, according to the Media Lab graph, the interaction energies between the pairs Welsh/Russian and Welsh/Bulgarian are negligible compared to the other interaction energies. The initial configuration, from Table A of [22], for this pair of parameters and these four languages is the following:

	$p_1$	$p_2$
$l_1$	+1	+1
$l_2$	-1	0
$l_3$	+1	+1
$l_4$	+1	-1

We ran a simulation of the dynamics of the system for two entailed parameters  $p_1, p_2$ , with the Hamiltonian described in Appendix A.

The dynamics depends on two parameters, the temperature  $T$  and the coupling energy  $E$  of the entailment relation. We ran simulations for high/low temperature (HT/LT) and high/low entailment energy (HE/LE).

In the first case, for the graph with languages  $\{l_1, l_2, l_3\} = \{\text{English, Russian, Bulgarian}\}$  and the parameters  $\{p_1, p_2\} = \{\text{Partial Definiteness, Definiteness Checking}\}$ , we consider an initial state as given in the first table above. The average value of spin for the two parameters is illustrated in Figures 5 and 6, in the different regimes of high temperature and high entailment energy (HT/HE), high temperature and low entailment energy (HT/LE), low temperature and high entailment energy (LT/HE), low temperature and low entailment energy (LT/LE). The final equilibrium states for the dynamics, in each of these cases, would then be as follows:

$(p_1, p_2)$	HT/HE	HT/LE	LT/HE	LT/LE
$l_1$	(+1,+1)	(+1,+1)	(-1,0)	(+1,+1)
$l_2$	(+1,0)	(+1,+1)	(-1,0)	(+1,+1)
$l_3$	(-1,0)	(-1,+1)	(-1,0)	(+1,+1)

In the second case, with languages  $\{l_1, l_2, l_3, l_4\} = \{\text{English, Welsh, Russian, Bulgarian}\}$  and with parameters  $\{p_1, p_2\} = \{\text{Strong Deixis, Strong Anaphoricity}\}$ , the initial state is given by the values of the parameters in the second table above, and the average value of spin in the different HT/HE, HT/LE, LT/HE, LT/LE regimes is illustrated in Figures 7 and 8. The final equilibrium states for the dynamics are then of the form

$(p_1, p_2)$	HT/HE	HT/LE	LT/HE	LT/LE
$l_1$	(+1,0)	(+1,-1)	(+1,+1)	(+1,-1)
$l_2$	(+1,-1)	(-1,-1)	(+1,+1)	(+1,-1)
$l_3$	(-1,0)	(-1,+1)	(+1,+1)	(-1,0)
$l_4$	(+1,+1)	(-1,-1)	(+1,+1)	(-1,0)

These examples are only illustrative and not entirely realistic, because we have singled out a small portion of the language graph that exhibits an interesting configuration of entailed parameters in the initial state, and we have run the simulation neglecting the interactions with all the other languages in the rest of the larger language graphs, which will also affect the behavior of the system. These examples, however, are interesting because they show situations where a configuration of entailed parameters reaches an equilibrium state where parameters of the individual languages have undergone some changes, but have not always converged to a configuration where all the parameters are aligned. While in the first example one obtains complete alignment of all the parameters in the low temperature and low energy regime, in the second example, even in this range, parameters do not fully align. This shows that the presence of entailment between syntactic parameters can have a substantial effect on the dynamics that differs significantly in outcome from the case where one assumes an independence hypothesis on parameters.

The problem of parameter variation and acquisition is discussed in depth in [21]. The dynamical model illustrated here can be applied, in principle, to the complete set of languages, parameters and relations given in [22] or to the more extensive list in [21]. However, as observed in [34], when run over the entire set of data of [22], the dynamics becomes extremely slow, due to the many relations, and it is hard to see the convergence to an equilibrium. However, one can still obtain interesting insights on the dynamics and on the variation of parameters across language families by using a coding theory approach, as in [25], [34].

## 2.8 Comparison with other models

Our article [35] is not the first application of methods from Statistical Physics in Linguistics; spin variables were also used in [4] and [28]. Our approach differs from these since we consider nodes

of the graph to be languages and we aim at modeling the effect of language interaction on language change, while [4] and [28] aim at describing mechanisms of language acquisition.

More precisely, the model of [4] is also based on the Principles and Parameters approach. However, the statistical physics model they construct is significantly different from the one we describe in the present article, and is completely unrelated. They are mostly interested in modeling language acquisition. They model by a Gibbs state the probability with which a speaker selects sentences, so that the choice should follow euphonic considerations—how best a given sentence fits the prosodic patterns of the language. A Gibbs state is a probability distribution where energy levels are weighted according to a thermodynamic temperature parameter, so that, at low temperature, the probability is almost entirely concentrated on the ground state (lowest energy), while at a higher temperature higher energy levels also contribute to more significant probabilities. To this purpose, they model prosody as a potential, which defines the Hamiltonian (energy functional) for their Gibbs state. The grammar is then estimated by a maximum likelihood argument maximizing the Gibbs probability measure.

The application of statistical physics models, and in particular of the Ising spin glass model, considered in §13.6.2–13.6.3 of [28] is a model of language acquisition and language change, where the author considers a population of linguistic agents, corresponding to the vertices of the graph, with edges representing the interaction between individual agents. The model also assumes that there are two distinct languages in the population. The adoption of one or the other language by a given speaker corresponds to the two possible spin states  $\pm 1$  at that vertex in the spin glass system. The inverse temperature parameter of the Ising model is interpreted as the probability with which speakers of each language produce “cues” (in a cue-based learning algorithm). The behavior of the system then models a situation where, above a certain critical temperature, individual agents behave independently and both languages are equally represented in the population, while below a critical temperature one language becomes dominant.

We have presented our model here primarily as a model of language change through language interaction affected by a measure of bilingualism in the population. It is an interesting question, which we hope to return to in future work, to see whether this type of spin

glass model can also be seen as a model of language acquisition, after suitably modifying the setting. In particular, we would like to connect the type of coupled Ising/Potts models considered here to the dynamical Markov models of language acquisition developed in [29], see also §13 of [28]. In the context of a possible use as a model of language acquisition, some interesting linguistic questions regarding parameters include whether there are default values of (some) parameters that are spontaneously temporarily set either innately or very early in development and that are switched in the presence of evidence enforcing a different value, or else retain their default. Such a hypothesis may be suitable for testing within a spin glass approach to language acquisition by simulating dynamics under a range of statistically chosen initial conditions, with or without constraints on the values of certain subsets of parameters. We do not consider this type of application in the present article, but we hope to investigate it further in upcoming work.

To our knowledge, a model like ours, where the syntactic parameters themselves are treated as spin variables in a spin glass model, and interactions are modeled by bilingualism data, has never been previously considered.

## Appendix A: The Mathematical Model

In this section we present briefly the actual mathematical model we work with, in sufficiently explicit detail, so that the results can be checked and reproduced. In particular, while for independent parameters this is just a collection of uncoupled Ising models on the same underlying graph, when entailment relations between parameters are also introduced, we need a new type of spin glass model, which has not been previously considered in the physics literature, and which we describe here in detail. It is obtained by coupling Ising models and Potts models with  $q = 3$  on the same graph, with an interaction term at the vertices that encodes the entailment of parameters. The resulting model can be regarded as a generalization of Potts models with external magnetic fields (see [13]), to which it reduces when all but one of the syntactic parameters are “frozen” and only one is left dynamical.

The data of a classical spin glass model (Potts model) with external magnetic field consist of:

- a finite graph  $G$  with vertices  $v \in V(G)$  and edges  $e \in E(G)$
- a finite set  $A$  of possible spin states at a vertex, with  $\#A = q \geq 2$
- the set of possible states of the system  $\Sigma = \text{Maps}(V(G), A)$ , describing all possible configurations of spins at the vertices of the graph
- interaction energies  $J_e$  along the edges  $e \in E(G)$ .
- Hamiltonian of the system: for states  $S \in \Sigma$

$$(1) \quad H(S) = - \sum_{e \in E(G): \partial(e) = \{v, v'\}} J_e \delta_{S_v, S_{v'}} - \sum_{v \in V(G)} B_v S_v,$$

where  $B = (B_v)$  is an external magnetic field, and  $\delta_{S_v, S_{v'}} = 0$  if  $S_v \neq S_{v'}$  and  $\delta_{S_v, S_{v'}} = 1$  if  $S_v = S_{v'}$ .

The first term in the Hamiltonian measures the degree of alignment between nearby spins, favoring configurations with aligned spins, while the second term measures the alignment of spins with the direction of the external magnetic field. The partition function of the system is

$$(2) \quad Z_G(\beta) = \sum_{S \in \Sigma} \exp(-\beta H(S)),$$

where the variable  $\beta = 1/T$  is a thermodynamic parameter equal to an inverse temperature (with Boltzmann constant set equal to one). The associated Gibbs probability measure is

$$(3) \quad \mathbb{P}_{G, \beta}(S) = \frac{e^{-\beta H(S)}}{Z_G(\beta)}.$$

**Independent parameters:** The full specification of the configuration of the system in the single-particle (single-language) basis is then given by

$$(4) \quad |\vec{S}\rangle = \bigotimes_{p \in \mathcal{P}} |\vec{S}_p\rangle = \bigotimes_{p \in \mathcal{P}} \left( \bigotimes_{\ell \in \mathcal{L}} |S_{\ell, p}\rangle \right)$$

where  $S_{\ell, p} \in \{-1, +1\}$ ,  $\mathcal{P}$  is the set of all independent parameters, and  $\mathcal{L}$  is the set of all languages.

With this picture in mind, we turn to the evolution. In this tensor product basis, we can simulate the evolution of each of the independent

syntactic parameters, by endowing the system with an interacting Hamiltonian for each parameter:

$$(5) \quad H_p = - \sum_{\ell, \ell' \in \mathcal{L}} J_{\ell \ell'} S_{\ell, p} S_{\ell', p}.$$

This Hamiltonian is minimized by the alignment of adjacent spins, as  $J_{\ell \ell'} \geq 0 \forall \ell, \ell'$ . Here, when we refer to the “low” and “high” temperature regimes, we mean the range in which the ratio  $T/\langle J_{\ell \ell'} \rangle$  is very small or, respectively, very large. Note that this is not the most general form of spin glass models, where the Hamiltonian typically contains additional disorder terms. We are also not considering here the possible presence of terms playing the role of an external magnetic field  $B$ , as in (2). Such terms would correspond to overall external effects on language change that do not come from the interaction between languages.

Figure 1 suggests that the vast majority of languages currently possess this parameter in the activated form  $+1$ . Evolving this system at very low temperature ( $T=0.000001$ ), we want to know the local magnetization, or the thermal average of the spin value at each vertex. Since we are in the low temperature regime, we expect that the vertices will tend to orient their spins in the same direction (either all of the languages will tend to possess the parameter or will lack it). Marking vertices with  $\langle S_{l,p} \rangle_H > 0$  light gray and vertices with  $\langle S_{l,p} \rangle_H < 0$  dark gray, we can look at a graph of these languages at equilibrium, given in Figure 2.

## 2.8 Parameters with entailment relations

We focus on a case with only two parameters  $p_1, p_2$  with an entailment relation as described in §2.5. Let  $S_{l,p}$  be the spin variables, as before, where  $S_{l,p_1} \in \{\pm 1\}$  and  $S_{l,p_2} \in \{\pm 1, 0\}$ . We consider a change of variable for the first spin, with the new variable  $X_{l,p_1} \in \{0, 1\}$  defined by  $S_{l,p_1} = \exp(\pi i X_{l,p_1})$ . This just corresponds to the two possible choices representing the group  $Z/2Z$  in additive form  $\{0, 1\}$  or multiplicative form  $\{+1, -1\}$ . We also associate to the spin variables  $S_{l,p_2} \in \{+1, 0, -1\}$  a variable  $Y_{l,p_2} = |S_{l,p_2}| \in \{0, 1\}$ . We then consider a Hamiltonian given by the sum of two terms  $H = H_E + H_V$ , where the term  $H_E$  expresses the edge interactions between different languages,

as above, written as

$$(6) \quad H_E = H_{p_1} + H_{p_2} = - \sum_{\ell, \ell' \in \text{languages}} J_{\ell \ell'} \left( \delta_{S_{\ell, p_1}, S_{\ell', p_1}} + \delta_{S_{\ell, p_2}, S_{\ell', p_2}} \right).$$

This term is the same as in the case of two independent parameters. The interaction between the parameters that accounts for the entailment relation is encoded in the  $H_V$  term as

$$(7) \quad H_V = \sum_{\ell} H_{V, \ell} = \sum_{\ell} J_{\ell} \delta_{X_{\ell, p_1}, Y_{\ell, p_2}}.$$

For  $J_l > 0$ , this gives an anti-ferromagnetic pairing between the variables  $X_{l, p_1}$  and  $Y_{l, p_2}$ .

The preferred energy states for the Hamiltonian  $H_V$  are those where either  $X_{l, p_1} = 0$  and  $Y_{l, p_2} = 1$ , or  $X_{l, p_1} = 1$  and  $Y_{l, p_2} = 0$ , which correspond, respectively, to the entailment relations  $S_{l, p_1} = 1$  and  $S_{l, p_2} = \pm 1$ , or  $S_{l, p_1} = -1$  and  $S_{l, p_2} = 0$ . Notice that the size of the parameters  $J_l$  in the term  $H_V$  in the Hamiltonian determine how strongly enforced the entailment relation is in a given language: for  $J_l \rightarrow \infty$ , an infinite energy barrier separates the ground states  $H_{V, l} = 0$  from the excited states  $H_{V, l} = J_l$ , hence the entailment is strongly enforced, while lower values of  $J_l$  would imply that the language  $l$  admits a certain frequency of exception to this syntactic rule. These should be interpreted in the same sense as the probabilistic approach to setting syntactic parameters discussed in [20], see also the discussion in §1.2.3 and §2.3 above. Notice that, if we freeze one of the parameter  $p_1$  and only consider the evolution of the second parameter, then the Hamiltonian above can be regarded simply as a case of Potts model with external magnetic field.

## 2.9 Metropolis–Hastings algorithm

The algorithm (see, for instance, §2 and §3 of [27]) is based on the detailed balance condition, which describes the stationary distribution of a Markov process, with given transition probabilities. The uniqueness of the stationary distribution is ensured by an ergodicity property of the Markov process, about which we comment more in detail below. Under these assumptions, the Metropolis–Hastings algorithm provides a construction of a Markov process with the desired properties, where the transition probabilities are obtained

in terms of acceptance-rejection rates of a possible transition of the system. Namely, given a current state of the system, one first provides a method of constructing a new proposed state. In the case where the graph is a lattice this is usually achieved by considering a single-spin-flip dynamics, where the new state differs from the old one by the flipping of a single spin, with a uniform probability. The new state is then automatically accepted if its energy (measured by the Hamiltonian of the spin glass model) is lower than or equal to that of the old state, while if the energy is higher, it is accepted with a probability that depends on the energy gap and a thermodynamic inverse temperature variable. This allows the dynamics of the system to escape from local minima (false vacua) and continue its descent towards the true energy minimizing equilibrium state.

The detailed balance condition (see [27]) is given by

$$(8) \quad \mathbb{P}(s)\mathbb{P}(s \rightarrow s') = \mathbb{P}(s')\mathbb{P}(s' \rightarrow s),$$

where  $P(s \rightarrow s')$  is the Markov process probability of transitioning from a state  $s$  to a state  $s'$ . The acceptance-rejection rates are given by

$$(9) \quad \mathbb{P}(s \rightarrow s') = \pi_A(s \rightarrow s') \cdot \pi(s \rightarrow s'),$$

where  $\pi(s \rightarrow s')$  is the conditional probability of proposing a state  $s'$  given the state  $s$  and  $\pi_A(s \rightarrow s')$  is the conditional probability of accepting the proposed state  $s'$  given  $s$ . The  $P(s \rightarrow s')$  are normalized to add up to one, by requiring that, in the remaining cases the proposed state is  $s'=s$ . The Metropolis–Hastings acceptance distribution is

$$(10) \quad \pi_A(s \rightarrow s') = \min\left\{1, \frac{\mathbb{P}(s') \pi(s' \rightarrow s)}{\mathbb{P}(s) \pi(s \rightarrow s')}\right\}$$

which satisfies the balance condition (8), so that

$$\frac{\pi_A(s \rightarrow s')}{\pi_A(s' \rightarrow s)} = \frac{\mathbb{P}(s') \pi(s' \rightarrow s)}{\mathbb{P}(s) \pi(s \rightarrow s')}.$$

The selection probabilities  $\pi(s \rightarrow s')$  are chosen so that ergodicity condition holds, for instance by single-spin-flip dynamics on a lattice. In particular, one can take  $P$  the Gibbs measure of a spin glass model, with acceptance probabilities  $\pi_A(s \rightarrow s')$  satisfying (8) given by

$$(11) \pi_A(s \rightarrow s') = \begin{cases} 1 & \text{if } H(s') - H(s) \leq 0 \\ \exp(-\beta(H(s') - H(s))) & \text{if } H(s') - H(s) > 0. \end{cases}$$

A new state  $s'$  with energy lower than or equal to that of  $s$  is automatically accepted, while one with a higher energy is accepted with probability  $\exp(-\beta(H(s') - H(s)))$ . See §3.1 of [27] for more details.

### 3.0 Ergodicity and Metropolis–Hastings algorithm

The ergodicity property is necessary for the working of the Metropolis–Hastings algorithm. A Markov chain is ergodic if and only if it is irreducible and aperiodic. Irreducibility is the condition that the Markov process can reach any possible state of the system starting from any given state in a finite number of steps. A state is periodic if the chain can return to it only at multiples of some period. Aperiodicity is the absence of periodic states. Under the assumption of single-spin-flip dynamics, ergodicity is satisfied for the Ising model on a lattice (see §3.1 of [27]). The graph we consider here is not a lattice. However, it makes sense in terms of the linguistic interpretation of the model, to make a similar type of simplifying assumption as in the single-spin-flip dynamics, namely that change happens one parameter at a time in one language at a time. This assumption may not be entirely realistic in the case of entailed parameters where the flipping of one parameter can force the change of others, but as we discussed previously, in our model entailment relations are enforced in a probabilistic way depending on a “coupling energy” between entailed parameters, so we can assume that the change of entailed parameters will not be simultaneous but subsequent to the change of one of them and enforced with a certain probability. For similar arguments in favor of single-spin-flip dynamics in models of language acquisition and language change, see [28]. Ergodicity of the Metropolis–Hastings algorithm, with the single-spin-flip dynamics, still holds for our graph.

We show here that the ergodicity condition for the Metropolis–Hastings algorithm holds in our setting. As we mentioned, ergodicity depends on the properties of irreducibility and aperiodicity. Irreducibility can be checked using the notion of accessibility. Given states  $s$  and  $s'$ , let  $\tau$  denote the minimum number of steps needed for the Markov chain to reach  $s'$  starting at  $s$ . The state  $s'$  is accessible from  $s$  if there is a positive probability that  $\tau < \infty$ . Suppose that the states  $s$  and

$s'$  differ at a certain set  $A \subset V(G)$  of sites, with  $k = \#A$  and  $N = \#V(G)$ . Then  $P(\tau < \infty) \geq P(\tau = k)$ . Let  $L_A$  be the set of all  $k!$  orderings of the elements of  $A$ . For  $j = 1, \dots, k!$ , and  $\sigma_j \in L_A$ , let  $s_j = (s_{j,r})_{r=0, \dots, k-1}$  be the corresponding list of  $k$  states in  $\Sigma$ , each obtained from the previous one by flipping the spin located at the site  $\sigma_j(r) \in A$ . The probability  $P(\tau = k)$  is then the sum of the products of the probability of obtaining a string  $\sigma_j$  in  $L_A$  by choosing uniformly randomly a sequence of  $k$  elements in  $V(G)$ , times the product of the probabilities of the flipping of the individual spins at the sites  $\sigma_j(r)$  for  $r = 1, \dots, k$ ,

$$\mathbb{P}(\tau = k) = \frac{1}{N!} \sum_{j=1}^{k!} \prod_{r=0}^{k-1} \pi_A(s_{j,r} \rightarrow s_{j,r+1}).$$

The precise values of these probabilities depends on the topology of the graph  $G$ , through the set of edges, that determine the change in energy between successive states, according to the acceptance probabilities of the Gibbs measure described above. It is however clear that  $\mathbb{P}(\tau = k) > 0$ , hence all states are accessible from a given one, which is equivalent to the irreducibility condition for the Markov chain.

For a single parameter, the set  $\Sigma_p$  of possible states has  $\#\Sigma_p = 2^{\#V(G)}$ . The number of states  $S'_p$  that are accessible from a given state  $S_p$ , in a single-spin-flip dynamics, is the number of states that differ from  $S_p$  by flipping a single spin at one of the vertices. Hence, for any given  $S_p$ , there are  $N = \#V(G)$  accessible states  $S'_p$ . The probability of choosing one of these states is uniform  $\pi(S_p \rightarrow S'_p) = 1/N$ . For a given state  $S_p$ , we have  $\mathbb{P}(S_p \rightarrow S_p) = 0$  if and only if all the accessible states  $S'_p$  with  $\pi(S_p \rightarrow S'_p) = 1/N > 0$  have lower energy, so that all  $\pi_A(S_p \rightarrow S'_p) = 1$  and  $\mathbb{P}(S_p \rightarrow S_p) = 1 - \sum_{S'_p} \mathbb{P}(S_p \rightarrow S'_p) = 1 - (1/N)\#\{S'_p\} = 0$ . Thus, states  $S$  with  $\mathbb{P}(S_p \rightarrow S_p) = 0$  are maxima of the energy. All other states  $S_p$  have  $\mathbb{P}(S_p \rightarrow S_p) > 0$ , hence they are aperiodic states of the Markov chain. If a Markov chain is irreducible, then all states have the same period, hence the remaining states must also be aperiodic. Together with irreducibility this implies ergodicity.

### 3.1 Independent parameters

If we focus on the behavior of a single syntactic parameter, our model is simply an Ising model on our graph of languages, with the given interaction energies, and we apply to it the Metropolis–Hastings algorithm as described above. This generates a discretized form of

gradient descent to the energy minima. When we consider more than one parameter at the same time, under the simplifying assumption that different syntactic parameters are independent of each other, we just run the same kind of Metropolis–Hastings algorithm on each parameter, by each time flipping a single spin variable of a single syntactic parameter. Namely, we proceed, as described above, using the Metropolis–Hastings algorithm with single-spin-flip dynamics. We propagate a state by proposing a new configuration and accepting it or rejecting it (and keeping the old one) with some probability. As discussed above, we assume the algorithm acts by trying to flip the spin at each vertex independently, although in reality this is more likely to occur in clusters. At each step, the algorithm chooses a vertex at random and tries to flip its spin, accepting or rejecting the flip according to the energy difference, according to the acceptance probabilities of the Gibbs measure. (This part of the algorithm ensures that the dynamics continues to descend towards the true energy minima, escaping false local minima.) Propagating the state for many steps produces a distribution of configurations that asymptotically approaches the thermal equilibrium probability density function (the Gibbs measure of equation (3)), allowing one to calculate thermal averages of various functions (such as magnetization) by evaluating them on the state at each step and averaging over steps.

The behavior of the average value of spin over vertices, in the case of a parameter with no entailment relations, is illustrated in Figure 4.

We reported here the case of the Subject-Verb parameter as an illustrative example: we described the behavior at low and high temperature  $T$ , averaging over a million steps. This is enough to obtain a clear understanding of the phase diagrams of simulations for all of the other parameters, when considered independently. The code in the GitHub repository can be used to generate analogous simulations for all parameters.

As expected, at low temperatures, since the system is close to  $T=0$ , it is also close to the ground state. In fact, a configuration randomly sampled from the equilibrium ensemble will also look like Figure 2 nearly 100% of the time, because of how dominant the all-up configuration is: it is nearly impossible at very low temperature for the state to propagate to a configuration that is not the ground state. At  $T=0$ , this is the only configuration in the ensemble. At this low temperature, the languages quickly converge during the evolution of the system to a state in which they all acquire the parameter in the

activated form +1. A more physical way to view this is to look at a plot of the average spin as a function of time, as given in the first graph of Figure 4. One may wonder why this system always ends up in this “all-up” state instead of an “all-down” state. The key is that the initial configuration is dominated by spin-up vertices, so the energy barrier is much higher for a droplet of spin-down vertices to expand and dominate the system than it is for the sea of spin-up vertices to swallow the droplets.

What happens when  $T$  is very large? Taking  $T = 20$ , the local magnetizations approach zero, with exactly half of the vertices approaching zero from the positive direction as in Figure 3. In fact, the local magnetizations vary with mean  $-2.3927 \times 10^{-4}$  and median  $7 \times 10^{-6}$  on the interval between  $-0.0186$  and  $0.0147$ . In this case, a configuration picked at random from the equilibrium ensemble is most likely to have approximately half of the vertices with spin up. Because  $T$  is large, a state can propagate to almost any other configuration fairly easily. At equilibrium, the languages here change approximately independently of each other, so the average (over vertices) spin fluctuates about zero, as in the second graph of Figure 4.

### 3.1 Metropolis–Hastings for entailed parameters

Several syntactic parameters without any entailment relations simply behave like independent uncoupled Ising models over the same underlying graph, with the same interaction energies but with different initial conditions, hence the same Metropolis–Hastings algorithm can be applied to each parameter. In the case of entailed parameters, that can also take an “undefined” value 0, in addition to the spin up and spin down values, we modify the algorithm accordingly, as we explain here, based on the Glauber dynamics for Potts models, [2], [11], so that we can still run the simulation accordingly.

In the case of entailed parameters, we run the analog of the Metropolis–Hastings algorithm for the Hamiltonian  $H = H_E + H_V$  on the graph of languages and language interactions. Since we now have spin variables with more than two possible states, instead of the single-spin-flip dynamics, one uses the appropriate Glauber dynamics. For general results about ergodicity and convergence time of Glauber dynamics on some classes of graphs, we refer the reader to [2], [11]. In the simple cases we analyzed in §2.7.1, we worked with a complete graph, so the results of [11] apply.

In this simulation, for the parameters with ternary values  $\{-1, 0, +1\}$ , we replace the previous acceptance probabilities with

$$(12) \quad \pi_A(s \rightarrow s \pm 1 \pmod{3}) = \begin{cases} 1 & \text{if } \Delta_H \leq 0 \\ \exp(-\beta \Delta_H) & \text{if } \Delta_H > 0. \end{cases}$$

where

$$\Delta_H := \min\{H(s + 1 \pmod{3}), H(s - 1 \pmod{3})\} - H(s).$$

### 3.2 Temperature and parameters

As discussed in §1.2.3, one can consider syntactic parameters not as frozen on either the up or down positions, with binary  $\{0, 1\}$  values (or  $\pm 1$  in multiplicative notation), but as a probability distribution  $\{P, 1 - P\}$  for some  $0 \leq P \leq 1$  expressing the frequency with which that parameter is expressed in the given language, see [20]. This corresponds to introducing a certain amount of “noise” in the setting of the parameter. A similar effect is obtained in a spin glass model through the temperature variable  $T$  (or the inverse temperature parameter  $\beta = 1/T$ ); while at low temperature spins tend to be frozen in the up or down position, at higher temperature they are free to fluctuate randomly between the positions. To relate these two phenomena, we can regard each language  $l$ , and each parameter  $p$ , as endowed with the additional data of probability distributions  $\{P_{l,p}, 1 - P_{l,p}\}$ , which represent how “noisy” the setting of the parameter  $p$  is in the language  $l$ .

In general, one expects that the probability distributions  $\{P_{l,p}, 1 - P_{l,p}\}$  will be different for different languages  $l$  and different parameters  $p$ . However, we can assume that they all depend on an overall thermodynamic parameter  $\beta$  that can be tuned to increase or decrease the amount of noise. In first approximation, one can assume for simplicity that all probabilities are the same, with a form like  $P_{l,p}(\beta) = P(\beta) = 1 - \frac{\exp(-\beta)}{2}$ , so that, at zero temperature ( $\beta = \infty$ ) the parameter  $p$  is set to  $+1$  without any noisiness (or to  $-1$  in the symmetric case), while for  $T \rightarrow \infty$  (at  $\beta = 0$ ) one has the maximum amount of noise, with the uniform distribution  $P = 1/2$  assigning equal probability to the two  $\pm 1$  values of the syntactic parameter. One can consider more general functions  $P_{l,p}(\beta)$ , which are continuous and

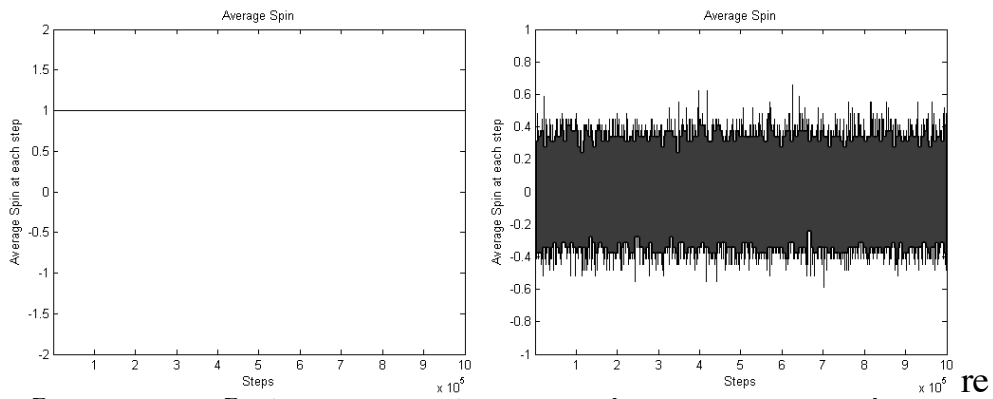
monotonically interpolating between the values  $P_{l,p}(\infty) = 1$  and  $P_{l,p}(0) = 1/2$ . For example, one can slightly modify the case discussed above by taking  $P_{l,p}(\beta) = 1 - \frac{\exp(-\beta\gamma_{l,p})}{2}$ , where the exponents  $\gamma_{l,p}$  should be fitted to data like those collected in [20], so that there is some common value  $\beta_0$  at which the probability distribution  $P_{l,p}(\beta_0)$  match the statistics for specific syntactic parameters in specific languages. At present, we do not yet have a sufficiently large set of data, of the type collected in [20], to fully implement this analysis, so we only outline here the general approach and we postpone the numerical implementation to future work.

## Appendix B: Plots of the Simulation Results

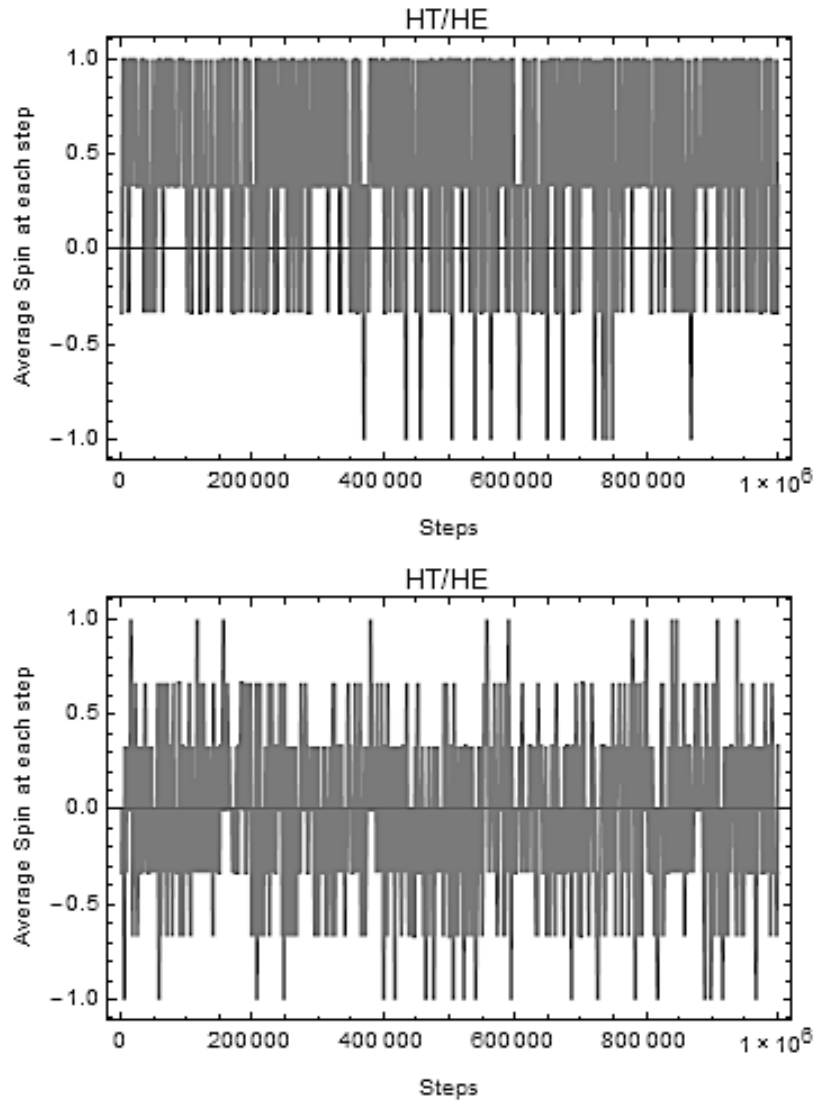
We collect here the graphics with the plots of the various simulation results that are described in the text.

Figure 4 shows the dynamics for independent parameters (no entailment) in the low temperature and high temperature regimes, while the following figures show the dynamics for a pair of parameters with entailment relations, in the various regimes with high/low temperature and high/low entailment energy.

Figure 5 shows the average value of spin for Partial Definiteness (left) and for Definiteness Checking (right) in the high temperature/high energy (HT/HE) regime; and in the low temperature/low energy (LT/HE) regime, as a function of the number of steps in the Monte Carlo simulation. Figure 6 shows the average value of Partial Definiteness (left) and Definiteness Checking (right) in the high temperature/low energy (HT/LE) regime, and in the low temperature/low energy (LT/LE) regime, as a function of the number of steps. Figure 7 shows the average value of spin for Strong Deixis (left) and Strong Anaphoricity (right) in the HT/HE regime and in the LT/HE regime, while Figure 8 shows the average value for the same parameters in the HT/LE regime and in the LT/LE regime.



regime  $T = 0.000001$  (left) and in the high temperature range  $T = 20$  (right), as a function of the number of steps in the Monte Carlo simulation.



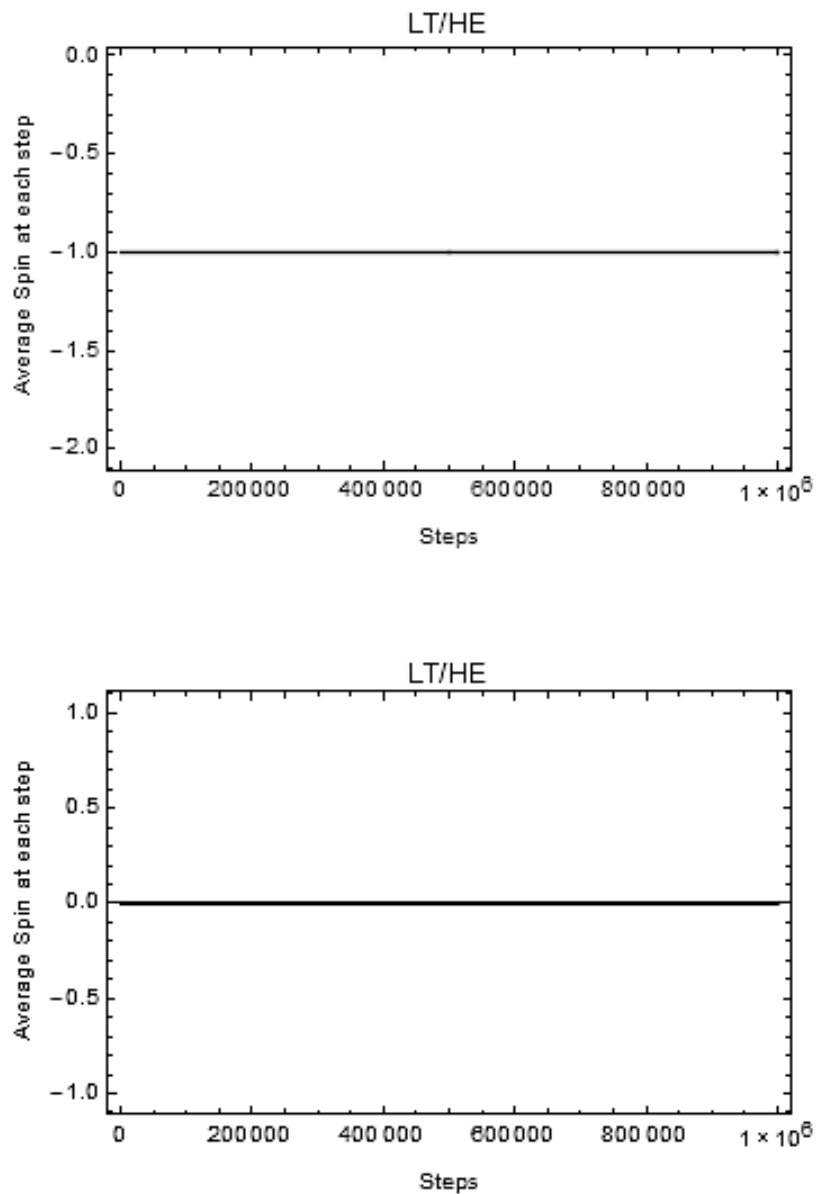
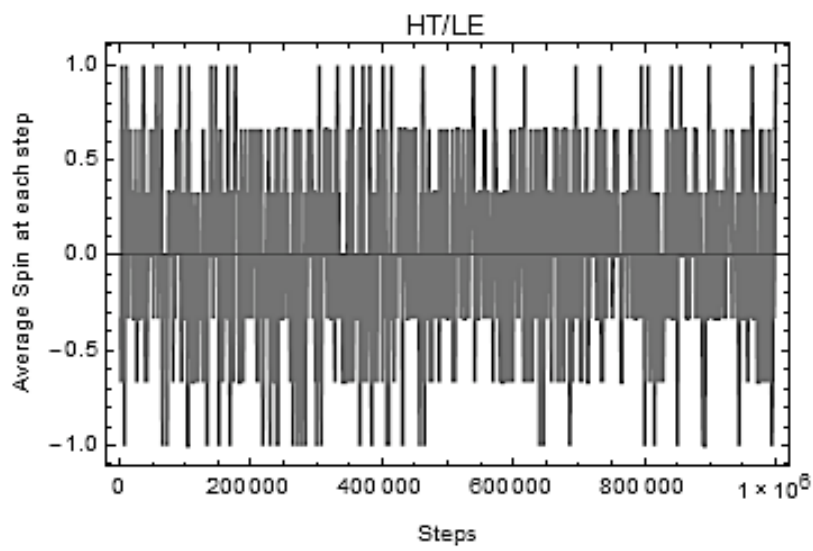
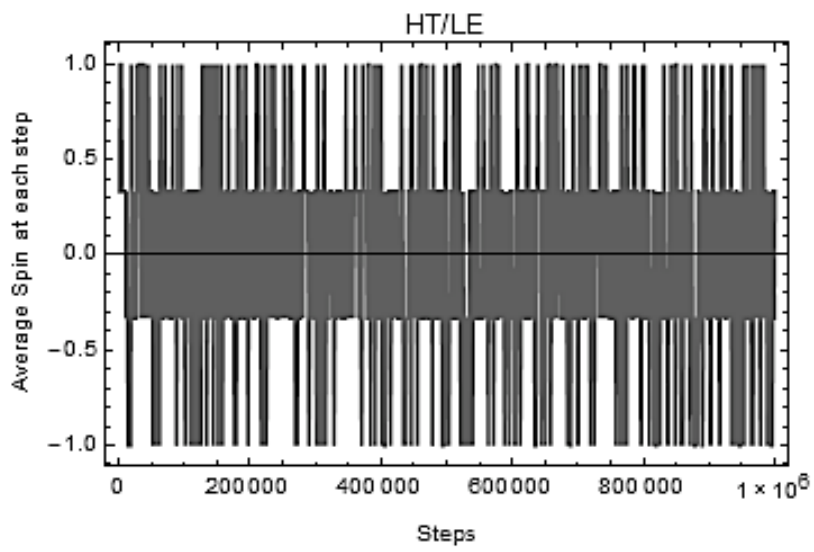


Figure 5: Average value of spin for  $p_1$  = Partial Definiteness (previous page, above), and for  $p_2$  = Definiteness Checking (previous page, below) in the high temperature/high energy (HT/HE) regime (this page, top) and in the low temperature/high energy (LT/HE) regime (this page, below), as a function of the number of steps in the Monte Carlo simulation.



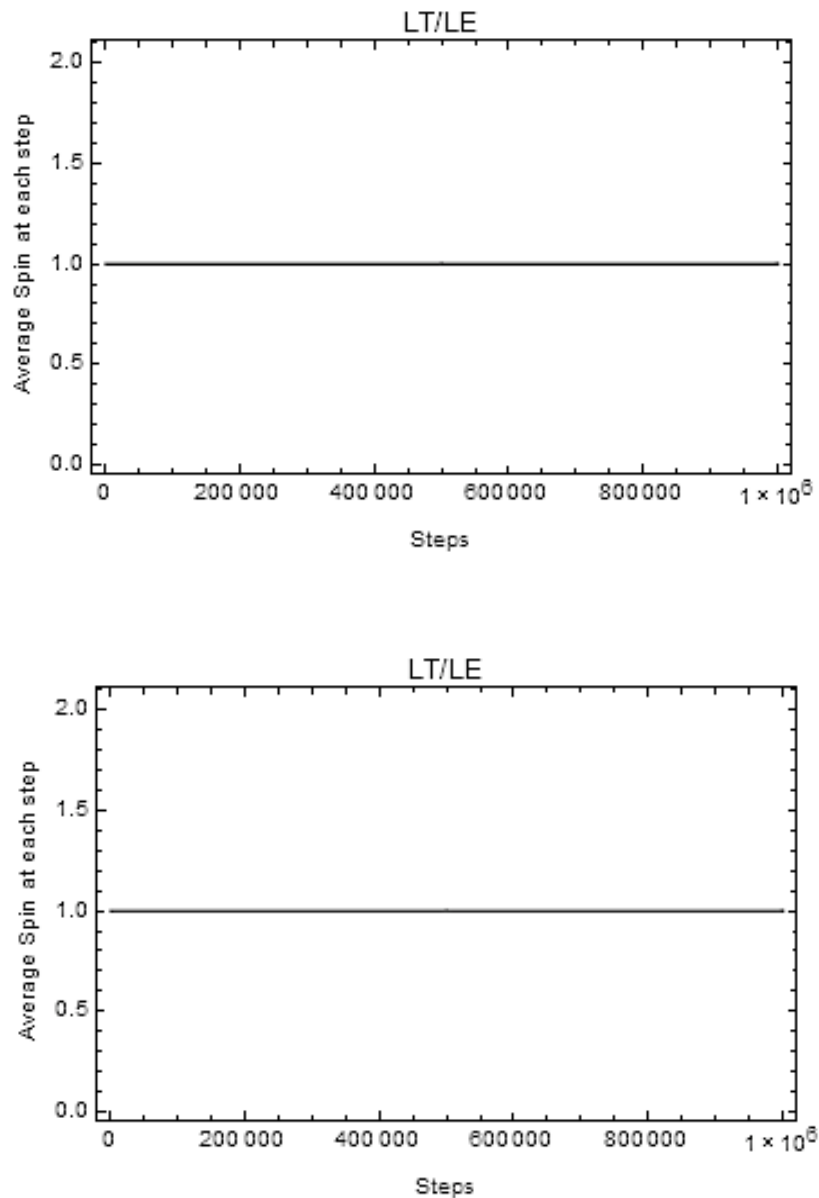
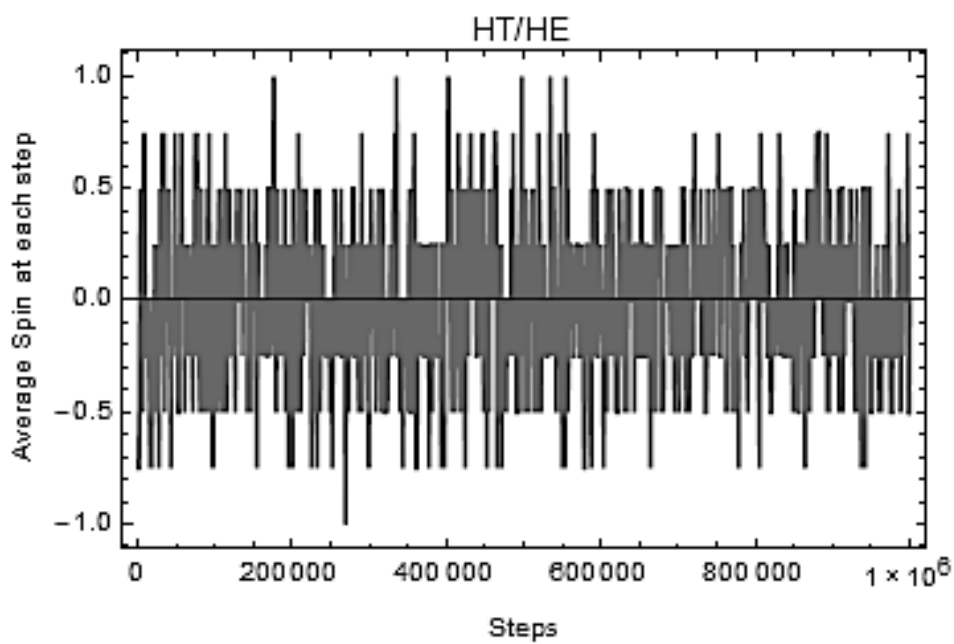
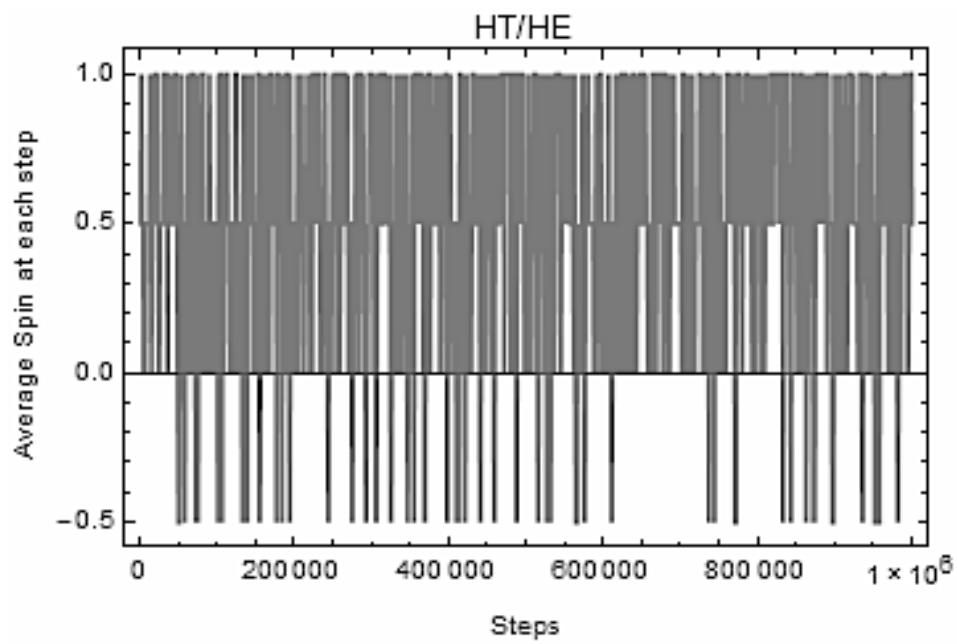


Figure 6. Average value of spin for  $p_1$  = Partial Definiteness (previous page, above) and for  $p_2$  = Definiteness Checking (previous page, below) in the high temperature/low energy (HT/LE) regime (this page, above) and in the low temperature/low energy (LT/LE) regime (this page, below), as a function of the number of steps.



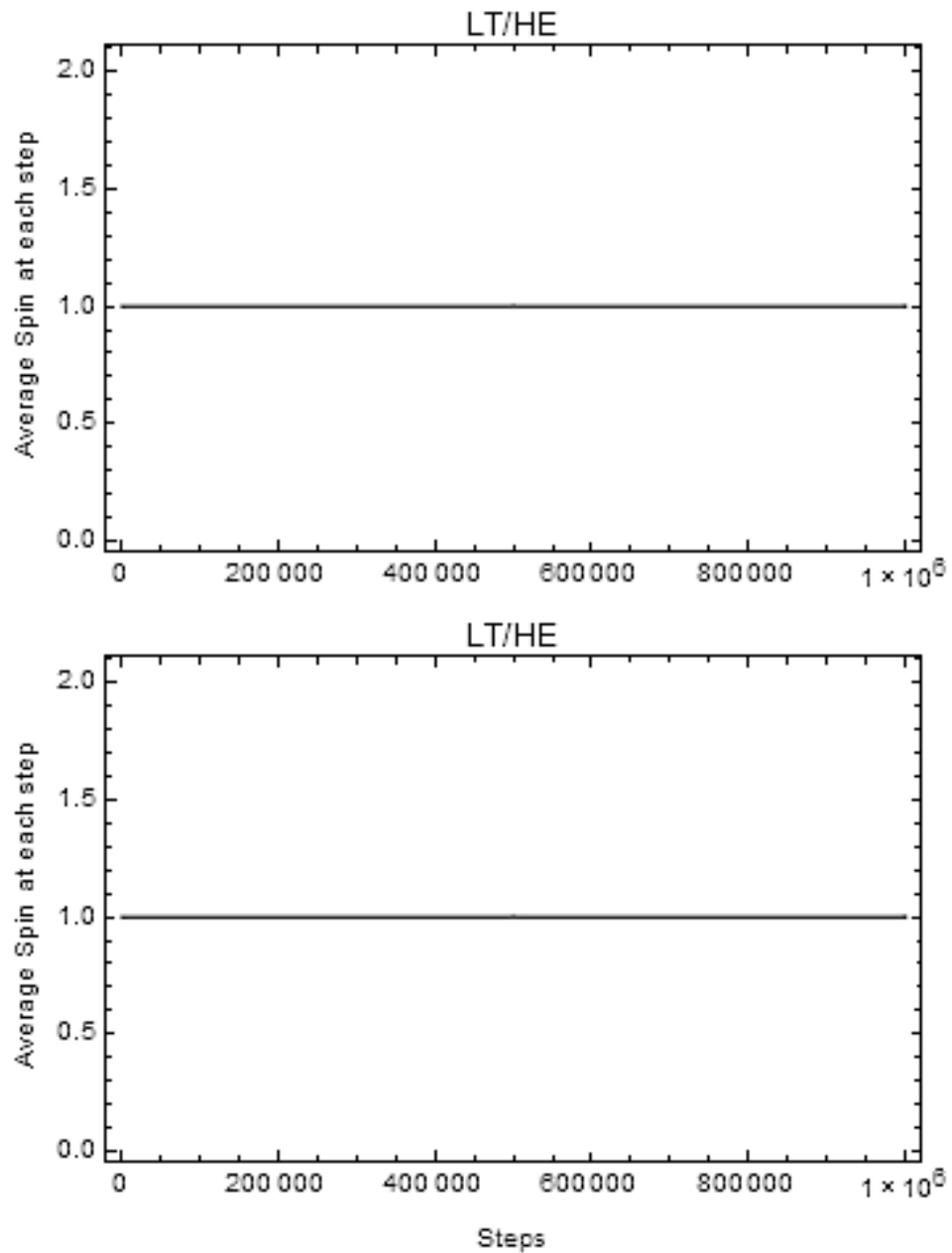
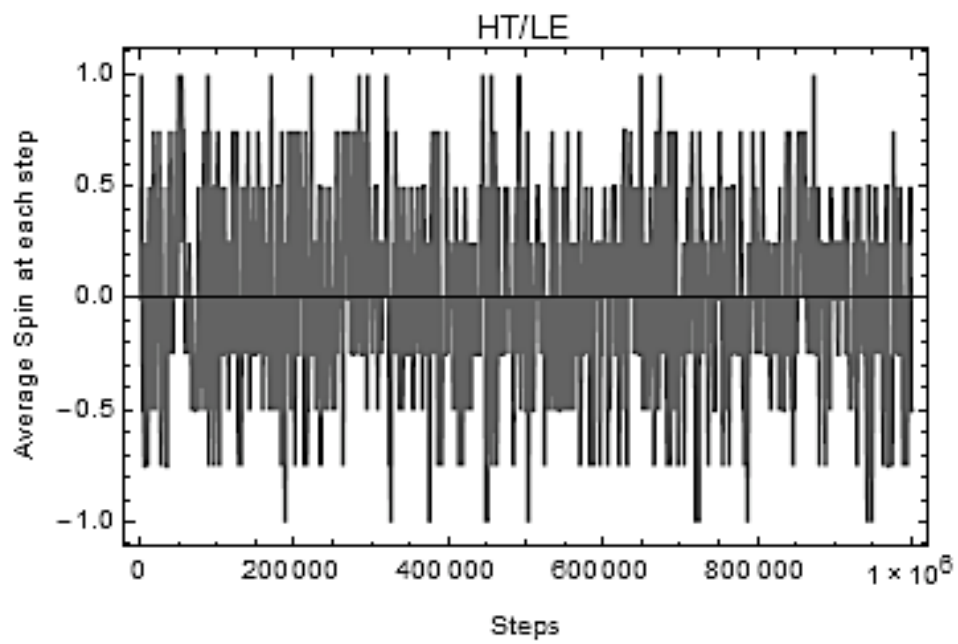
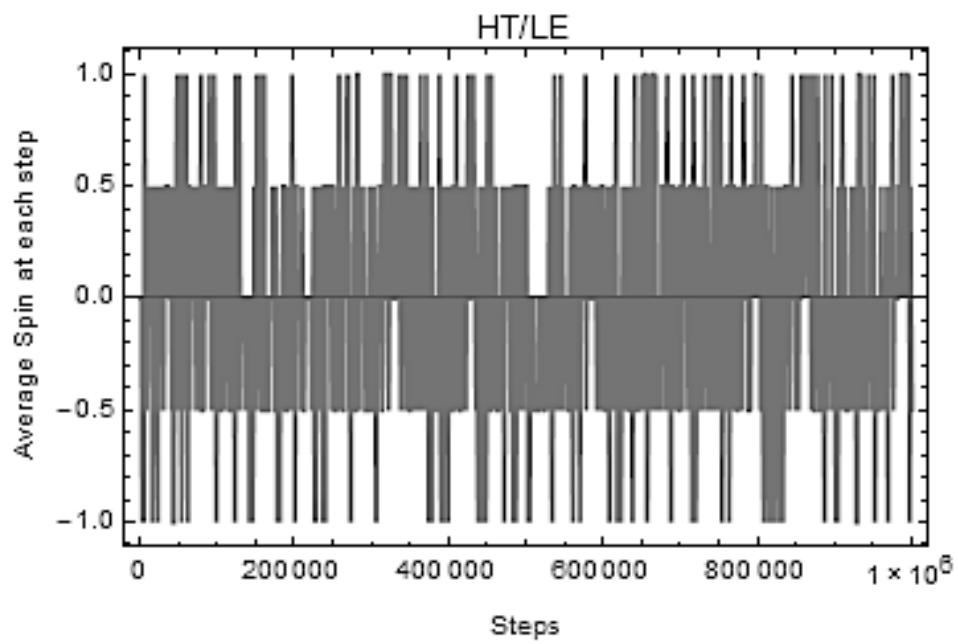


Figure 7. Average value of spin for  $p_1 = \text{Strong Deixis}$  (previous page, above) and for  $p_2 = \text{Strong Anaphoricity}$  (previous page, below) in the HT/HE regime (this page, top) and the LT/HE regime (this page, below), as a function of the number of steps.



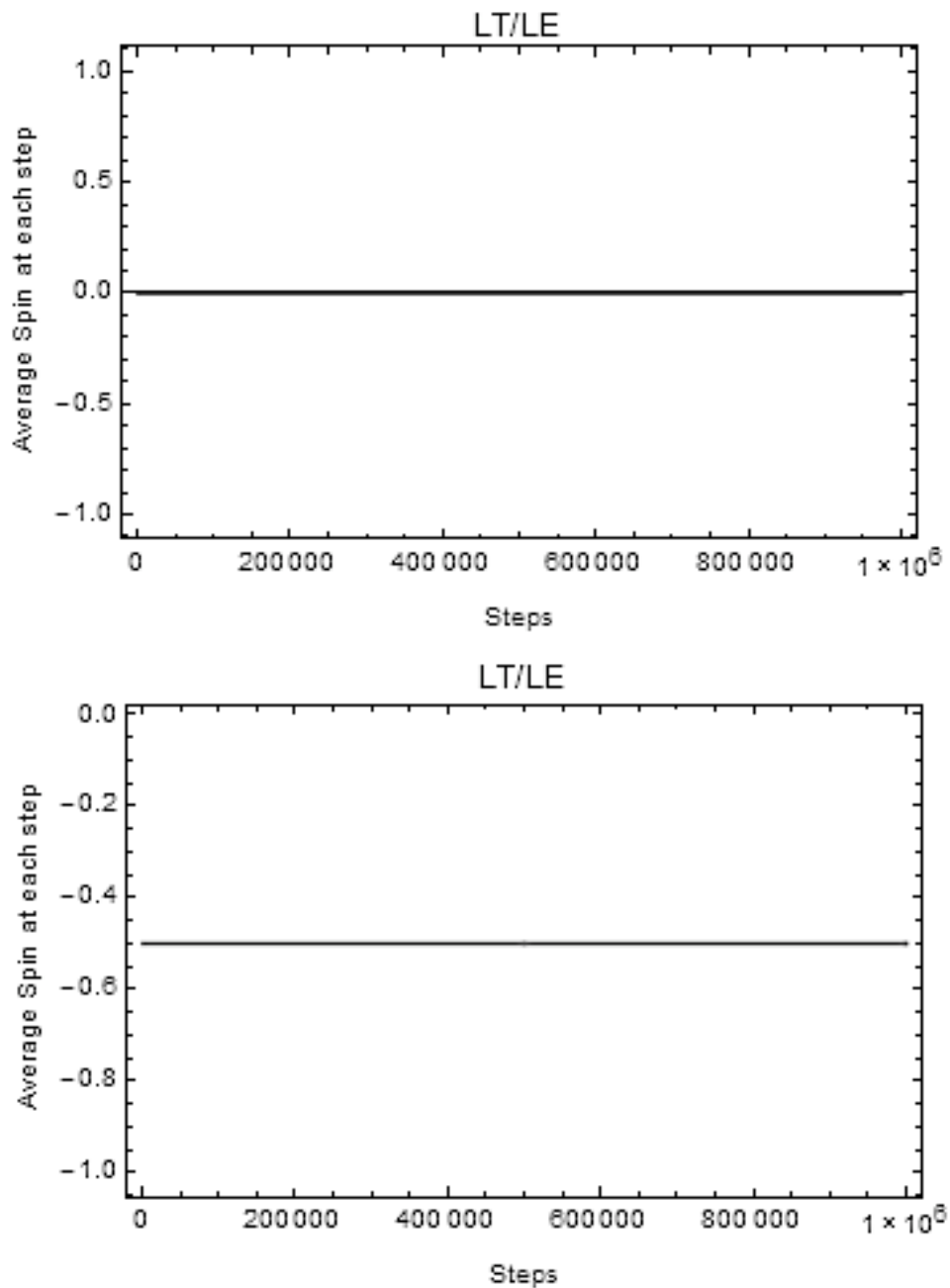


Figure 8. Average value of spin for  $p_1 =$  Strong Deixis (previous page, top) and for  $p_2 =$  Strong Anaphoricity (previous page, bottom) in the HT/LE regime (this page, top) and the LT/LE regime (this page below), as a function of the number of steps.

## Acknowledgment

This work was performed as part of the activities of the last author's Mathematical and Computational Linguistics lab and CS101/Ma191 class at Caltech. The last author was partially supported by NSF grants DMS-1201512, DMS-1707882, and PHY-1205440. We thank Massimo Piattelli-Palmarini for many helpful comments.

## Works Cited

1. Baker, M. 2001. *The Atoms of Language*. New York: Basic Books.
2. Berger, N., C. Kenyon, E. Mossel, Y. Peres. 2005. Glauber dynamics on trees and hyperbolic graphs. *Probability Theory and Related Fields* 131: 311–340.
3. Berwick, R.C., P. Ho. 2013. Languages Do Not Show Lineage-Specific Trends in Word-Order Universals. Talk given at LSA.
4. Cassandro, M., P. Collet, A. Galves, C. Galves. 1997. A Statistical-Physics Approach to Language Acquisition and Language Change. *Physica A: Statistical Mechanics and its Applications* 263: 427–437.
5. Chomsky, N. 1982. *Lectures on Government and Binding*. Dordrecht: Foris Publications.
6. Chomsky, N., H. Lasnik. 1993. The Theory of Principles and Parameters. In *Syntax: An International Handbook of Contemporary Research*. Amsterdam: de Gruyter, 506–569.
7. Clark, R., I. Roberts. 1993. A Computational Approach to Language Learnability and Language Change. *Linguistic Inquiry* 24: 299–345.
8. Cole, P. 1974. Indefiniteness and Anaphoricity. *Language* 50: 665–674.
9. Cornilescu, A., A. Nicolae. 2011. On the Syntax of Romanian Definite Phrases: Changes in the Patterns of Definiteness Checking. In *The Noun Phrase in Romance and Germanic*. Amsterdam: Benjamins, 193–221.
10. Croft, W. 2008. Evolutionary Linguistics. *Annual Review of Anthropology* 37: 219–234.
11. Cuff, P., J. Ding, O. Loidor, E. Lubetzky, Y. Peres, A. Sly. 2012. Glauber Dynamics for the Mean-Field Potts Model. *Journal of Statistical Physics* 149: 432–477.

12. Dahl, Ö. 1999. Does Adaptation Really Help Us to Explain Language Change? *Zeitschrift für Sprachwissenschaft* 18: 209–211.
13. Ellis-Monaghan, J.A., I. Moffatt. 2011. The Tutte-Potts Connection in the Presence of an External Magnetic Field. *Advances in Applied Mathematics* 47: 772–782.
14. Fischer, K.H., J.A. Hertz. 1991. *Spin Glasses*. Cambridge: Cambridge University Press.
15. Galves, C. (Ed.). 2012. *Parameter Theory and Linguistic Change*. Oxford: Oxford University Press.
16. Haspelmath, M., 2008. Parametric Versus Functional Explanations of Syntactic Universals. In *The Limits of Syntactic Variation*. Amsterdam: John Benjamins, 75–107.
17. Itkonen, E. 1999. 1999. Functionalism Yes, Biologism No. *Zeitschrift für Sprachwissenschaft* 18: 219–221.
18. Krauth, W. 2006. *Statistical Mechanics: Algorithms and Computations*. Oxford: Oxford University Press.
19. Lenz, F. 2003. *Deictic Conceptualisation of Space, Time, and Person*. Amsterdam: Benjamins.
20. Liu, H. 2010. Dependency Direction as a Means of Word-Order Typology: A Method Based on Dependency Treebanks. *Lingua* 120: 1567–1578.
21. Longobardi, G. 2017. Principles, Parameters, and Schemata. A Constructivist UG. *Linguistic Analysis* 41:3-4 (in this volume).
22. Longobardi, G., C. Guardiano. 2009. Evidence for Syntax as a Signal of Historical Relatedness. *Lingua* 119: 1679–1706.
23. Longobardi, G., C. Guardiano, G. Silvestri, A. Boattini, A. Ceolin, 2013. Towards a Syntactic Phylogeny of Modern Indo-European Languages. *Journal of Historical Linguistics* 3: 122–152.
24. MacSwan, J. 2014. *Grammatical Theory and Bilingual Codeswitching*. Boston: MIT Press.
25. Marcolli, M. 2016. Syntactic Parameters and a Coding Theory Perspective on Entropy and Complexity of Language Families. *Entropy* 18: 110.
26. Nerbonne, J. 2010. Measuring the Diffusion of Linguistic Change, *Philosophical Transactions B Royal Society* 365: 3821–3828.
27. Newman, M.E.J., G.T. Barkema. 1999. *Monte Carlo Methods in Statistical Physics*. Oxford: Oxford University Press.
28. Niyogi, P. 2006. *The Computational Nature of Language Learning and Evolution*. Boston: MIT Press.
29. Niyogi, P., R.C. Berwick. 1996. A Language Learning Model for Finite Parameter Spaces. *Cognition* 61: 161–193.

30. Ronen S., B. Goncalves, K.Z. Hu, A. Vespignani, S. Pinker, C.A. Hidalgo. 2014. Links That Speak: The Global Language Network and its Association with Global Fame. *Proceedings of the National Academy of Sciences* (PNAS) 111: 5616–5622.
31. Park, J.J., R. Boettcher, A. Zhao, A. Mun, K. Yuh, V. Kumar, M. Marcolli. Prevalence and Recoverability of Syntactic Parameters in Sparse Distributed Memories. arXiv:1510.06342 [cs.CL].
32. Rizzi, L. 2016. On the Format and Locus of Parameters: The Role of Morphosyntactic Features. *Linguistic Analysis* 41:3-4 (in this volume).
33. Rozanov, Y.A. 2012. *Markov Random Fields*. Berlin: Springer.
34. Shu, K., M. Marcolli. Syntactic Structures and Code Parameters, arXiv:1610.00311.
35. Siva, K., J. Tao, M. Marcolli. Spin Glass Models of Syntax and Language Evolution. arXiv:1508.00504 [cs.CL].
36. Sokal, A.D. 2005. The Multivariate Tutte Polynomial (Alias Potts Model) for Graphs and Matroids. In *Surveys in Combinatorics 2005*. London Mathematical Society Lecture Note Series 327: 173–226. Cambridge: Cambridge University Press.
37. Stein, D.L., C.M. Newman. 2013. *Spin Glasses and Complexity*. Princeton: Princeton University Press.
38. Taylor, A. 1994. The Change from SOV to SVO in Ancient Greek. *Language Variation and Change* 6: 1–37.
39. Woolford, E. 1983. Bilingual Code-Switching and Syntactic Theory. *Linguistic Inquiry* 14: 520–536.
40. SSWL Database of Syntactic Structures of World Languages: <http://sswl.railsplayground.net/>
41. TerraLing database <http://www.terraling.com/>