Journal of Mathematical Biology manuscript No.
(will be inserted by the editor)

W. Garrett Mitchener

Bifurcation Analysis of the Fully Symmetric Language Dynamical Equation (preprint)

October 22, 2002

Abstract. In this paper, I study a continuous dynamical system that describes language acquisition and communication in a group of individuals. Children inherit from their parents a mechanism to learn their language. This mechanism is constrained by a universal grammar which specifies a restricted set of candidate languages. Language acquisition is not error-free. Children may or may not succeed in acquiring exactly the language of their parents. Individuals talk to each other, and successful communication contributes to biological (or cultural) fitness. I provide a full bifurcation analysis of the case where the parameters are chosen to yield a highly symmetric dynamical system. Populations approach either an incoherent steady state, where many different candidate languages are represented in the population, or a coherent steady state, where the majority of the population speaks a single language. The main result of the paper is a description of how learning reliability affects the stability of these two kinds of equilibria. I rigorously find all fixed points, determine their stabilities, and prove that all populations tend to some fixed point. I also demonstrate that the fixed point representing an incoherent steady state becomes unstable in an $\mathcal{S}_n$-symmetric transcritical bifurcation as learning becomes more reliable.

1. Introduction

Human languages consist of two parts: a lexicon, which is a set of words and their meanings, and a grammar, which is a set of rules for assembling and interpreting sentences. Children acquire their native language by hearing example sentences from their parents through which they learn both the lexicon and the grammar [16]. The general problem of acquiring grammar only from example sentences is known to be impossible without constraints on the rules of the grammar [9]. A widely accepted theory is that humans have a built-in set of constraints known as *universal grammar* or *UG* which guides the acquisition of native languages [3, 20]. UG operates even when

W. Garrett Mitchener: Program in Applied and Computational Mathematics, Fine Hall, Washington Road, Princeton, NJ 08544-1000. e-mail: `wmitchen @ princeton.edu`

Key words: Language – Evolution – Dynamical System – Bifurcation – Symmetry

Mathematics Subject Classification (1991): 91F20, 92D15, 37N25

the input is exceptionally impoverished, as in the cases of creolization [2] and the spontaneous invention of sign languages [21]. Children in these situations develop a fully functional grammar despite the lack of grammatical input and speak or sign quite differently from their parents. Some aspects of grammar, such as the word order, seem to be represented in the brain as a finite number of parameters with a small number of possible settings [4], and learning these parameter settings is equivalent to choosing among a finite number of possible classes of grammars [8]. The grammar acquisition process is not completely understood, but many theories are based on the idea that children set parameters based on specific cues from the sample sentences they hear, and many changes in grammars over time may be explained by a change in the linguistic environment that triggers a change in one of these parameters in the next generation [15].

A number of mathematical frameworks have been proposed for modeling the evolution of languages [13, 14, 17–19]. This paper is concerned with the model described in [13], in which Komarova, Niyogi and Nowak use evolutionary principles to model a population where each member speaks one language and benefits from being able to communicate with the rest of the population; this paper extends the analysis in [13] and adds several new results. We assume that the members of the population have a common lexicon and that certain lexical aspects of grammar, such as the forms of pronouns and tense morphemes, are fixed, leaving each child with a choice among a finite number of grammars. Children are assumed to learn their native language by hearing their parents speak, but the language acquisition process is subject to error, so they may end up with a grammar different from that of their parents. Learning error often results from ambiguity in the sample sentences children hear: The presence of a foreign language or multiple dialects can create enough linguistic noise that children are unable to determine exactly which grammatical rules to adopt [15, 16]. Rather than try to model the acquisition process in detail, this paper will treat learning in an abstract manner and deal only with the probability of making a learning error. The dynamical system which models this population is called the *language dynamical equation.* The focus of this paper is to provide a complete bifurcation analysis of the language dynamical equation in a special case where the parameters are chosen to make the dynamical system highly symmetric.

Two classes of population states are of primary interest. A coherent population is one in which the majority of members speak one language, and an incoherent population is one in which many languages are spoken by a significant fraction of the population. The tension between learning error and selection influences whether a given initial population reaches equilibrium in a coherent or incoherent state. The language dynamical equation contains selection terms which drive the population toward coherence, and mutation terms, corresponding to imperfect learning, which drive the population toward incoherence. If children are very likely to make mistakes in acquiring their language, then all languages can be equally distributed in the population, and the selection terms which give people a benefit for their ability to communicate have little effect. When children learn reliably, a language which is already widespread tends to become even more popular. Parents who speak it will almost surely pass it on to their children, and the selection term will be high for that language because its speakers can communicate perfectly with each other, and they form a large fraction of the population. When learning is very unreliable, the only stable equilibrium is an incoherent state. As the parameters of the model change to reflect increased learning reliability, stable coherent equilibria appear. The incoherent equilibrium eventually becomes unstable, and almost all populations tend to a coherent equilibrium. The bifurcation analysis presented here provides a mathematical description of how this transition from incoherence to coherence takes place.

Section 2 sets up the language dynamical equation and describes how it generalizes the replicator equation and the quasispecies equation.

In its fully general form, the language dynamical equation is a system of non-linear ordinary differential equations in an arbitrary number of dimensions, and a complete analysis of such a system is probably not possible. However, a considerable amount of information can be derived from a special case of the model in which the parameters are set to make the different grammars completely interchangeable. Section 3 describes these parameter settings.

The resulting system of ODEs has permutation symmetry and can be analyzed in detail. The fixed-point analysis here adds detail to the results in [13]. Section 4 gives an outline of the bifurcation scenario and pictures from the three-grammar case. In Section 5, we determine the locations of all fixed points and the parameter values for which they exist. Section 6 describes the linear stability analysis of all fixed points. Bifurcations occur when the parameters are such that the linearization of the system is singular at a fixed point. All such bifurcations of fixed points are found in Section 7, including the $\mathcal{S}_n$ transcritical bifurcation in which the incoherent equilibrium reverses stability.

Further analysis in Section 8 shows that the symmetric language dynamical equation happens to be nearly a gradient system, and a number of results about gradient systems can be adapted and applied to it. With a few short arguments, we will rule out closed orbits, homoclinic loops, and directed heteroclinic cycles. Finally, we show that all populations tend to some fixed point.

2. The language dynamical equation

Consider a large population of freely interacting individuals with identical language faculties. We assume that they share a lexicon, and each individual uses one of a finite number n of different grammars $G_1, \ldots, G_n$ in speaking and understanding sentences. The population as a whole is analogous to a quasispecies, because the members have a lot in common, namely the ability to use language, without being identical, as they have different grammars. It is assumed that all individuals interact with each other, and reproduce at a rate dependent upon some fitness measure of the grammar they use. Reproduction of language is accomplished by learning: Children learn the grammar of their parents by hearing example sentences, with the possibility that they might make mistakes. Learning mistakes can be thought of as mutations, as they cause parents speaking G_i to bear offspring speaking G_j. The population is represented by $x_1, \ldots, x_n$ where x_j is the fraction of people speaking G_j. We require $\sum x_j = 1$.

For this model, the relative fitness of an individual is based upon its grammar and the composition of the population. Given constants $A_{i,j}$ representing the probability that a sentence spoken at random from G_i can be parsed by a speaker of G_j, we define the fitness of G_i to be

$$F_i = \sum_{k=1}^{n} (\alpha A_{i,k} + (1-\alpha) A_{k,i})\, x_k. \tag{1}$$

That is, fitness depends on the ability for a speaker of G_i to be understood by and to understand a speaker of G_k. This is a measure of the similarity of the two grammars and is independent of the actual speakers. In F_i, the ability to communicate with G_k is weighted proportionally to its abundance x_k. If the parameter α is large, more benefit comes from being understood, and if it is small, more benefit comes from being able to understand. For the rest of this analysis, we give equal weight to both terms by setting $\alpha = \frac{1}{2}$ which yields

$$F_i = \sum_{k=1}^{n} \frac{A_{i,k} + A_{k,i}}{2} x_k. \tag{2}$$

In formulating the dynamics, we also need the variable ϕ representing the average fitness:

$$\phi = \sum_{k=1}^{n} F_k x_k. \tag{3}$$

Note that ϕ is a quadratic form in the x_j's.

To model learning, we define a row-stochastic matrix Q such that $Q_{i,j}$ is the probability that a teacher speaking G_i produces a student speaking G_j. The entries of Q are analogous to mutation rates. The Q matrix is row stochastic, meaning the sum of each row is 1, because every student must learn some grammar.

The *language dynamical equation* is an ODE representing the population dynamics:

$$\begin{aligned} \dot{x}_j &= \sum_{i=1}^{n} F_i x_i Q_{i,j} - \phi x_j \\ &= (F_j Q_{j,j} - \phi)\, x_j + \sum_{i \neq j} F_i x_i Q_{i,j}. \end{aligned} \tag{4}$$

Each G_i reproduces at a basic rate F_i, but a fraction $Q_{i,j}$ of the offspring erroneously learn G_j. The second form illustrates that the net reproductive rate of G_j depends on how much its fitness, scaled by learning reliability, exceeds the average fitness ϕ. This is the selection term. The other terms are mutation terms and represent contributions due to learning error. Note that this equation is cubic in the x_i variables. Only A and Q are constant in time; ϕ and F are functions of x.

Note that the total population $\sum x_j$ remains fixed at 1, because its time derivative is zero. All orbits of interest are therefore confined to an invariant hyperplane defined by $x \cdot \mathbf{1} = 1$, where $\mathbf{1}$ is a vector whose entries are all 1. Furthermore, if $x_j = 0$, then $\dot{x}_j \geq 0$ as it is a sum of terms each of which is at least 0. In particular, if $x_j(t_0) \geq 0$, it cannot at some later time cross the hyperplane perpendicular to the basis vector e_j because the vector field points the wrong way. Therefore, the positive orthant, defined as the subset of $\mathbf{R}^n$ where each $x_j \geq 0$, is a trapping region. The intersection of the invariant hyperplane and the positive orthant is a simplex S_n. For example, S_3 is an equilateral triangle, and S_4 is a regular tetrahedron.

The language dynamical equation combines ideas from both the replicator equation [12] and the continuous quasispecies equation [5,6]. It builds on the basic structure of the replicator equation, but adds mutation as in the quasispecies equation, and the reproductive rates are dependent upon the structure of the population.

3. Parameter settings for permutation symmetry

The fully general model (4) is too complex to analyze without some simplifying assumptions. Following Komarova *et al.* [13], we will constrain the A and Q matrices so that there are only two free parameters and the system as a whole exhibits permutation symmetry, that is, all the grammars will be interchangeable. With these constraints, we can analyze the dynamical system thoroughly despite its non-linearity.

For the rest of the paper, we will assume the following form for A and Q:

$$A = \begin{pmatrix} 1 & a & \cdots & a \\ a & 1 & \cdots & a \\ \vdots & \vdots & \ddots & \vdots \\ a & a & \cdots & 1 \end{pmatrix}, \tag{5}$$

$$Q = \begin{pmatrix} q & u & \cdots & u \\ u & q & \cdots & u \\ \vdots & \vdots & \ddots & \vdots \\ u & u & \cdots & q \end{pmatrix}, \quad \text{where } u = \frac{1-q}{n-1}. \tag{6}$$

The parameters a and q now completely determine the model. All off-diagonal entries of A are the same, so the probability that two people who use different grammars understand each other is the same no matter which grammars they use. Children acquire their grammar without error with probability q and mistakenly acquire each other grammar with probability u.

For convenience, we define variables M_k representing the k-th moment of the vector x:

$$M_k = \sum_{j=1}^{n} x_j^k. \tag{7}$$

Simplifying the original form of the language dynamical equation (4) and incorporating the restrictions on A and Q yields the following expression for the dynamics:

$$\dot{x}_j = (1-a)\left((q-u)x_j^2 + uM_2 - x_j M_2\right) - aunx_j + au. \tag{8}$$

Note that this vector field has the permutation group on n letters, commonly denoted $\mathcal{S}_n$, as its symmetry group, as all variables x_j are interchangeable. We will refer to (8) as the *fully symmetric language dynamical equation*, and the rest of the paper is concerned with this restricted form of (4).

4. Outline of the bifurcation scenario

To illustrate the bifurcation scenario for the fully symmetric language dynamical equation, we display here some pictures from the three-grammar case. They show the simplex as a triangle, where the corners represent the extreme values of (x_1, x_2, x_3), namely $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$. The parameter a is fixed at 0.5, and q varies.

For low values of q, the picture is as shown in Figure 1. There is a single fixed point which will be called the *uniform fixed point* in the middle of the simplex. It is a stable sink, meaning nearby populations tend to it in forward time. In this case, all populations tend to the uniform fixed point. It represents an incoherent population where each language is spoken in equal proportion. Here, the inaccuracies in learning drown out the effects of the selection terms in the model.

As q increases, a number of symmetric saddle-node bifurcations occur, resulting in Figure 2. In each corner of the simplex, a pair of fixed points appears, one stable sink close to the corner, and one unstable saddle between the sink and the uniform fixed point. The stable sinks in the corners represent coherent populations, where one language is spoken by a large portion of the population. Populations which start close to a corner move to a coherent state, and populations which start close to the center move to the uniform fixed point and incoherence. All the stable sinks have a basin of attraction, meaning a set of nearby population states which tend to them in forward time. The saddle points have only a thin manifold of population states which tend to them in forward time, and these stable manifolds form the boundaries between the basins of attraction of the sinks. In this situation, learning has become accurate enough that the population can choose a dominant language. When a large portion of the population speaks one language, the fitness term in the ODE for that language is high because those people understand each other perfectly. This causes the language to be spoken more widely in the future. However, populations still have a choice between coherence in the corners, and incoherence in the middle.

When q exceeds a particular value, the saddle points collide with the uniform fixed point in what is known as an $\mathcal{S}_n$-symmetric transcritical bifurcation. The result is shown in Figure 3. In this bifurcation, the uniform fixed point reverses its stability and becomes an unstable source. The saddle points pass through it and re-organize themselves, as their stable manifolds must now form boundaries between basins of attraction in the corners, but no longer in the middle. All populations (except the few on the stable manifolds of saddle points) now choose a dominant language and move toward one of the sinks in the corners. In this case, the inaccuracies of learning are drowned out by the selection term, and incoherence is no longer stable.

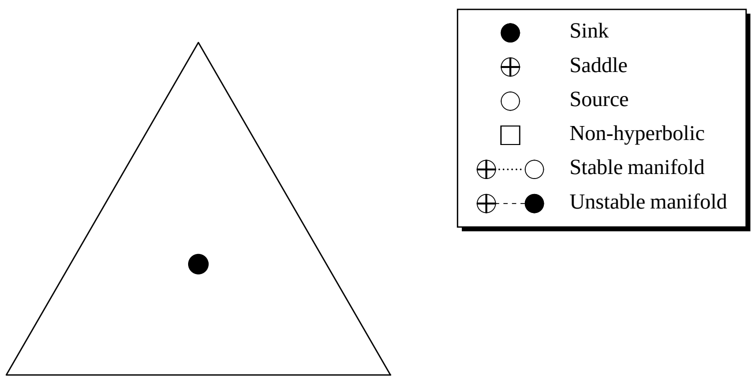

Fig. 1. Phase portrait with $a = 0.5$, $q = 0.85$.

In higher dimensions, the basins of attraction of the various sinks are more complex, and there are more saddle points which come into existence before the $\mathcal{S}_n$-symmetric transcritical bifurcation. The higher dimensional cases are hard to draw; however the three-language case drawn here should provide enough illustration to give the reader some intuition for the analysis that follows.

5. Locating the fixed points

We will now locate all the fixed points of the fully symmetric language dynamical equation, and identify the parameter ranges for which they exist. In particular, the order in which fixed points come into existence can be completely determined. It is reasonable to guess that the fixed points of (8) will have some symmetric form. In particular, we make the assumption

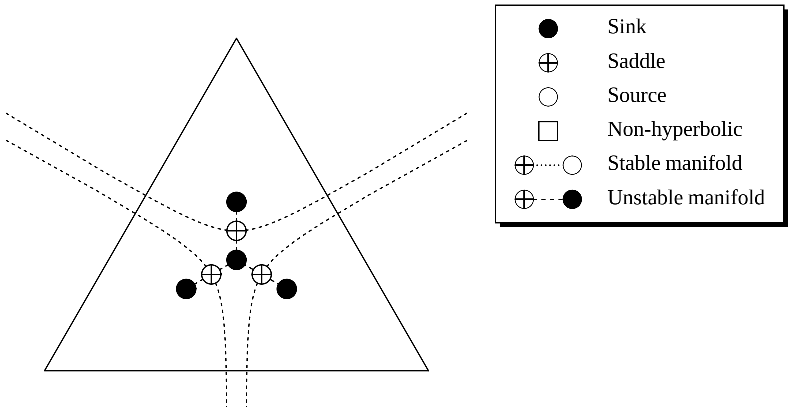

Fig. 2. Phase portrait with $a = 0.5$, $q = 0.8575$.

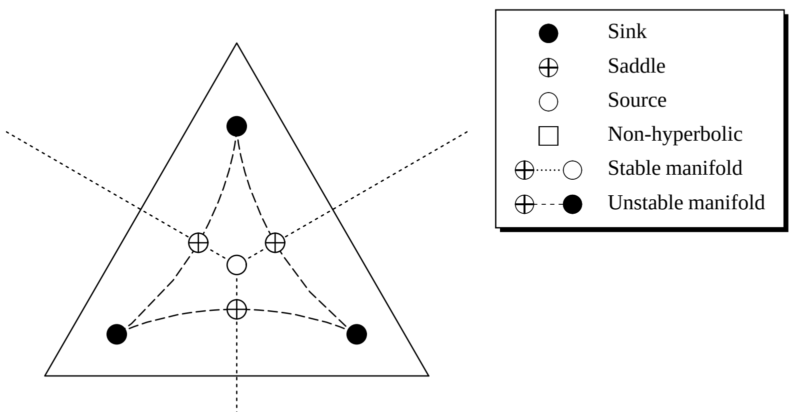

Fig. 3. Phase portrait with $a = 0.5$, $q = 0.9$.

that at fixed points, m grammars will share the majority of the population equally, and the rest will split the remainder equally.

Proposition 1. *Every fixed point $\bar{x}$ of (8) has m entries equal to some number Z and $n-m$ entries equal to $(1-mZ)/(n-m)$.*

Proof. Suppose $\bar{x}$ is a fixed point. At that point, M_2 is some constant which depends upon $\bar{x}$. Then each coordinate $\bar{x}_j$ must be a root of the polynomial

$$(1-a)\left((q-u)Z^2 + uM_2 - ZM_2\right) - aunZ + au = 0. \tag{9}$$

This polynomial, which comes from (8), is quadratic in Z, so it has at most two real roots. Therefore, each $\bar{x}_j$ is limited to be one of at most two values, and we may assume m of them are of one value and $n-m$ are of the other. Since $\sum \bar{x}_j = 1$, the fixed point must be of the required form. $\square$

We define $X^{(m)}$ and $Y^{(m)}$ to be the roots of (9), with $X^{(m)}$ referring to the larger. A fixed point with m entries equal to $X^{(m)}$ and $n-m$ entries equal to $Y^{(m)}$ will be called an *m-up fixed point*. There are $\binom{n}{m}$ ways to distribute m grammars of majority frequency $X^{(m)}$ and $m-n$ grammars of minority frequency $Y^{(m)}$ among the n entries of x, yielding $\binom{n}{m}$ symmetrical m-up fixed points.

The next step is to give explicit expressions for all of these fixed points, and determine the values of q for which they appear. We fix a, and consider what happens as q increases from $1/n$ to 1.

First, there is one fixed point corresponding to $m = 0$ or $m = n$ called the *uniform* solution. It is given by

$$x_j = \frac{1}{n}, \text{ where } j = 1 \dots n.$$

This fixed point represents a population where all grammars are spoken with equal frequency. It exists for all a and q, as can be seen by plugging it into (8). When solving for m-up fixed points, the uniform solution will always show up as an extra solution where $X^{(m)}$ and $Y^{(m)}$ are both $1/n$.

Other fixed points can be found by substituting the form described in Proposition 1 into (8). That is, we solve for the possible values of each x_j by setting

$$\begin{aligned} x_j &= Z, \\ M_2 &= mZ^2 + (n-m)\left(\frac{1-mZ}{n-m}\right)^2, \end{aligned}$$

which yields a cubic equation. It turns out that $Z = 1/n$ is always a root of this equation, which reflects the fact that the uniform solution is of the required form for every m. Extracting the factor of $(nZ-1)$ from the cubic yields the following quadratic:

$$\begin{aligned} &(a-1)m(n-1)Z^2 \\ &+ (a-1)(1+2m(q-1)-nq)Z \\ &- (a(1-m-n)-1)(q-1) = 0. \end{aligned} \tag{10}$$

The roots are found with the quadratic formula, yielding

$$Z_{\pm}^{(m)} = -\frac{1-2m+2mq-nq}{2m(n-1)} \pm \frac{\sqrt{d}}{2m(1-a)(n-1)}, \tag{11}$$

where the discriminant d is given by

$$\begin{aligned} d =&(1-a)\Big(4m(n-1)(1-q)(a+am-an-1) \\ &+ (1-a)(1-2m(1-q)-nq)^2\Big). \end{aligned} \tag{12}$$

The quadratic equation (10) was set up to look for values of Z such that some fixed point has m elements equal to Z. Therefore,

$$\begin{aligned} X^{(m)} &= Z_+^{(m)}, \\ Y^{(m)} &= Z_-^{(n-m)} = \frac{1-mZ_+^{(m)}}{n-m}. \end{aligned} \tag{13}$$

If q is small enough, d will be negative, and there will be no m-up fixed points. When q is such that $d = 0$, there will be some sort of saddle-node bifurcation, as the m-up and $(n-m)$-up fixed points will be identical. The bifurcation value of q may be found by solving the quadratic equation $d = 0$. The appropriate root is

$$\begin{aligned} \hat{q}_m = \frac{1}{(a-1)(n-2m)^2}\Big(&2m(n-m)(2+a(n-3)) + (a-1)n \\ &- 2(n-1)\sqrt{(1+a(m-1))(1+a(n-m-1))m(n-m)}\Big). \end{aligned} \tag{14}$$

Note that $\hat{q}_m = \hat{q}_{n-m}$, which implies that the m-up and $(n-m)$-up fixed points will appear at the same time as q increases. See Figure 4 for an example graph of $\hat{q}_m$. As can be seen from its concave-down shape, the m-up fixed points appear in a particular order: first the 1-up and $(n-1)$-up fixed points, then the 2-up and $(n-2)$-up, and so on.

More rigorously, we can make the substitution $m = n/2 + h$. After some simplification, (14) becomes

$$\hat{q}_m = \frac{2+a(n-3)}{2(1-a)} + \frac{(n-1)n\left(1-a+\frac{an}{2}\right)}{4(1-a)}g(h), \tag{15}$$

where

$$g(h) = \frac{\sqrt{(1-\beta_1 h^2)(1-\beta_2 h^2)}-1}{h^2},$$

and

$$\beta_1 = \frac{a^2}{\left(1-a+\frac{an}{2}\right)^2} \text{ and } \beta_2 = \frac{4}{n^2}.$$

Note that $\beta_1 = \beta_2$ only if $a = 1$, in which case all the languages are identical, or if $a = 1/(1-n) < 0$; neither case is of interest here, so $\beta_1 \neq \beta_2$. The important thing to notice is that $\hat{q}_m$ is a positive constant plus a positive constant times $g(h)$, and β_1 and β_2 are positive and unequal. Thus, we only need to establish the shape of the graph of $g(h)$ to determine the shape of the graph of $\hat{q}_m$.

Proposition 2. *The function $g(h)$ has a global maximum at $h = 0$ and is concave down for $-n/2 \leq h \leq /2$.*

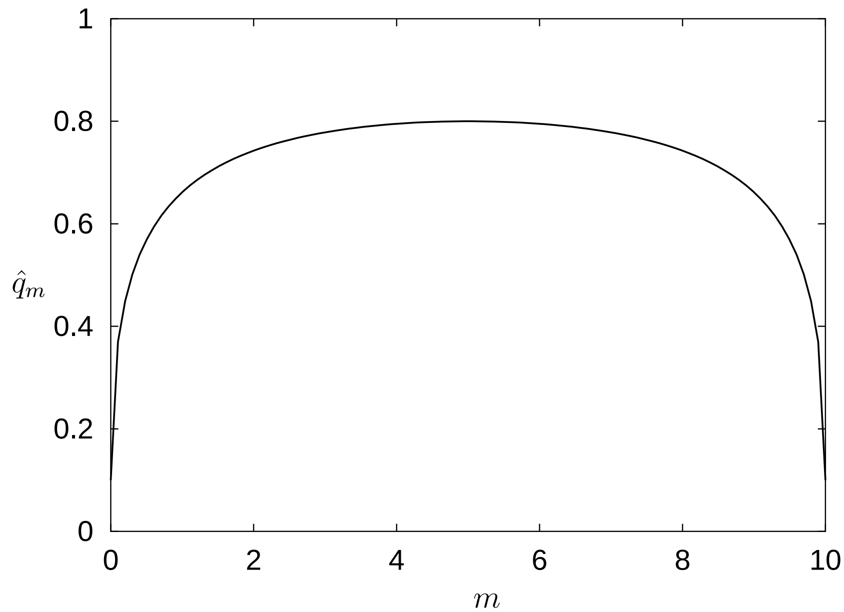

Fig. 4. Plot of $\hat{q}_m$, with $a = 0.2$ and $n = 10$. The m-up fixed points do not exist until $q > \hat{q}_m$.

Proof. Note that for small h, we can expand $\sqrt{1 - \beta h^2}$ into the Taylor series $1 + \beta h^2/2 + \beta h^4/8 + O(h^6)$, which quickly gives the expansion $g(h) = -(\beta_1 + \beta_2)/2 - (\beta_1 - \beta_2)^2 h^2/8 + O(h^4)$. From this series, we can read off

$$\begin{aligned} g(0) &= -\frac{\beta_1 + \beta_2}{2} < 0, \\ g'(0) &= 0, \\ g''(0) &= -\frac{(\beta_1 - \beta_2)^2}{4} < 0. \end{aligned}$$

This analysis proves that g has a critical point at $h = 0$, which is a local maximum by the second derivative test. In fact, this is the only critical point of g, and therefore a global maximum, as may be seen by analyzing its derivative directly. If g' is to be zero, its numerator must be zero, which implies, after some manipulation, that

$$\beta_1 \beta_2 h^4 = (\beta_1 + \beta_2)^2 h^4,$$

which has only the solution $h = 0$. Therefore g has a single critical point, a global maximum at $h = 0$, and is concave down everywhere else. $\square$

This lemma implies that $\hat{q}_m$, which is just a scaled and translated version of g, must always have the shape suggested by Figure 4. In particular, $\hat{q}_m$

has a global maximum at $m = n/2$, given by

$$\hat{q}_{max} = \hat{q}_m|_{m=\frac{n}{2}} = \frac{1 + n + a(n^2 - n - 1)}{n(2 - 2a + an)}, \tag{16}$$

after much simplification.

6. Linear stability analysis

Now that all the fixed points of the fully symmetric language dynamical equation have been found, their stabilities must be determined by linear stability analysis. In this section, we will compute the Jacobian matrix of the vector field in (8) at the various fixed points, derive expressions for its eigenvalues, and determine their multiplicities. In Section 7, we will determine the parameter values for which each is a source, a sink, or a saddle.

We will work with the n variables $x_1, \ldots, x_n$ and treat them as independent. The fact that the region of interest is a simplex embedded in an $(n-1)$-dimensional hyperplane will come into play after the n-by-n Jacobian has been computed. An alternative would be to replace x_n by $1 - (x_1 + \cdots + x_{n-1})$ and work in $n - 1$ independent variables, but that method yields results that are somewhat harder to visualize as the simplex is no longer easily visible.

The Jacobian matrix for (8) has entries of two types:

$$\frac{\partial \dot{x}_i}{\partial x_i} = (1-a)\left(2x_i\left(q - x_i\right) - M_2\right) - aun, \tag{17}$$

and for $j \neq i$:

$$\frac{\partial \dot{x}_i}{\partial x_j} = 2(1-a)\left(u - x_i\right) x_j. \tag{18}$$

For simplicity of notation in this section, j is assumed to be different from i whenever used as a subscript. Due to the symmetry of the ODE, the same expression is obtained for any $j \neq i$.

Since each x_i will have to be one of two values, the Jacobian matrix has a special structure which makes its eigenvalues relatively easy to find. In particular, define the following variables:

$$\begin{aligned}
c_1 &= \left.\frac{\partial \dot{x}_i}{\partial x_i}\right|_{x_i = X^{(m)}} & c_2 &= \left.\frac{\partial \dot{x}_i}{\partial x_i}\right|_{x_i = Y^{(m)}} \\
c_3 &= \left.\frac{\partial \dot{x}_i}{\partial x_j}\right|_{x_i = X^{(m)}, x_j = X^{(m)}} & c_4 &= \left.\frac{\partial \dot{x}_i}{\partial x_j}\right|_{x_i = X^{(m)}, x_j = Y^{(m)}} \\
c_5 &= \left.\frac{\partial \dot{x}_i}{\partial x_j}\right|_{x_i = Y^{(m)}, x_j = Y^{(m)}} & c_6 &= \left.\frac{\partial \dot{x}_i}{\partial x_j}\right|_{x_i = Y^{(m)}, x_j = X^{(m)}}
\end{aligned}$$

With the preceding definitions, the Jacobian of (8) at an m-up fixed point with the first m entries equal to $X^{(m)}$ takes the form

$$J = \left(\begin{array}{cccc|cccc} c_1 & c_3 & \cdots & c_3 & c_4 & c_4 & \cdots & c_4 \\ c_3 & c_1 & \cdots & c_3 & c_4 & c_4 & \cdots & c_4 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c_3 & c_3 & \cdots & c_1 & c_4 & c_4 & \cdots & c_4 \\ \hline c_6 & c_6 & \cdots & c_6 & c_2 & c_5 & \cdots & c_5 \\ c_6 & c_6 & \cdots & c_6 & c_5 & c_2 & \cdots & c_5 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c_6 & c_6 & \cdots & c_6 & c_5 & c_5 & \cdots & c_2 \end{array}\right). \tag{19}$$

The lines separate columns 1 to m and rows 1 to m from the rest. Due to the permutation symmetry of the dynamical system, the coordinates of any other m-up fixed point may be derived from this one by shuffling its entries; its Jacobian may be found by conjugating J with a permutation matrix, so it will have the same eigenvalues. Thus, to determine the stabilities of all fixed points, it is sufficient to analyze J.

In addition to the special form of J, a further observation makes it possible to quickly determine eigenvalues of J: We are interested in J only at fixed points within the simplex. Since the $(n-1)$-dimensional hyperplane containing the simplex is invariant, $n-1$ of the eigenvectors should lie within this hyperplane, and the last eigenvector should lie outside. The special form of J suggests that we try an eigenvector of the form

$$v = \begin{pmatrix} r \\ \vdots \\ r \\ s \\ \vdots \\ s \end{pmatrix}.$$

The first m entries are the same and therefore invariant under permutations of the first m variables. The last $n-m$ are similarly invariant. The equation $Jv = \lambda v$ reduces to the following two-dimensional eigenvalue problem:

$$\begin{aligned} c_1 r + (m-1)c_3 r + (n-m)c_4 s &= \lambda r, \\ m c_6 r + c_2 s + (n-m-1)c_5 s &= \lambda s. \end{aligned} \tag{20}$$

The assumption that v lies in the hyperplane of the simplex gives an additional equation, $v \cdot \mathbf{1} = 0$, which expands into

$$mr + (n-m)s = 0. \tag{21}$$

Using (21) to solve for s in terms of r and substituting that expression for s in the first equation of (20) yields

$$\lambda_1 = c_1 + (m-1)c_3 - mc_4. \tag{22}$$

A particular eigenvector v_1 corresponding to λ_1 may be found from (21), for example, by setting $r = (n - m)$ and $s = -m$.

A second eigenvalue may be determined by computing the trace of the system (20) and subtracting λ_1. The result is

$$\lambda_0 = c_2 + (n - m - 1)c_5 + mc_4. \tag{23}$$

However, the corresponding eigenvector v_0 points outside the simplex and is not of interest here.

The remaining $n - 2$ eigenvalues may be found by looking at subspaces orthogonal to v_1. In particular, the $m - 1$ vectors

$$-e_1 + e_k \text{ for } k = 2 \dots m$$

are eigenvectors such that

$$J(-e_1 + e_k) = (c_1 - c_3)(-e_1 + e_k).$$

The notation e_k means the k-th standard basis vector of $\mathbf{R}^n$. Likewise, the $n - m - 1$ vectors

$$-e_{m+1} + e_k \text{ for } k = m + 2 \dots n$$

are eigenvectors with

$$J(-e_{m+1} + e_k) = (c_2 - c_5)(-e_{m+1} + e_k).$$

In summary, if we assume $m > 0$, then λ_0 and $\lambda_1 = c_1 + (m-1)c_3 - mc_4$ are eigenvalues of multiplicity 1, $\lambda_2 = c_1 - c_3$ is an eigenvalue of multiplicity $m - 1$, and $\lambda_3 = c_2 - c_5$ is an eigenvalue of multiplicity $n - m - 1$. In the special case where $m = 0$, we get only two eigenvalues, λ_0 of multiplicity 1, and λ_3 of multiplicity $n - 1$.

7. Bifurcations of fixed points

In Section 6, we determined the eigenvalues of the linearized fully symmetric language dynamical equation at all fixed points. Bifurcations of fixed points can be detected by looking for parameter settings which cause these eigenvalues to equal zero. The parameter a is considered to be fixed, and q to vary from $1/n$ to 1. In this section, we determine what parameter values cause the eigenvalues to be zero and account for all bifurcations involving just fixed points. From this information, we can determine the signs of all the eigenvalues and therefore the stability of each fixed point. First, we handle the m-up fixed points, which come into existence through saddle-node bifurcations. Then, we discuss the uniform fixed point, which always exists, but undergoes a reversal of stability.

7.1. Bifurcations of the m-up fixed points

There are two special values of q corresponding to bifurcations: $\hat{q}_{max}$ which corresponds to a collision of many fixed points at the center of the simplex, and $\hat{q}_m$ which corresponds to several simultaneous saddle-node bifurcations in which the m-up and $(n-m)$-up fixed points come into existence. A number of tricks will be used to solve for these bifurcation points. To illustrate the technique, we first find the sign changes of λ_2 because it is the simplest of the three eigenvalues to work with and the calculations can be carried out by hand. The same calculations work for λ_3 and λ_1, but for λ_1 they become unwieldy and are best carried out with the aid of a computer algebra system.

Proposition 3. *For an m-up fixed point where $n > m > n/2$, the eigenvalue λ_2 is strictly negative for $q < \hat{q}_{max}$, zero for $q = \hat{q}_{max}$, and strictly positive for $q > \hat{q}_{max}$. If $1 < m \leq n/2$, then the eigenvalue $\lambda_2 \geq 0$ for $q = \hat{q}_m$ and strictly positive for $q > \hat{q}_m$.*

Proof. Since λ_2 is of multiplicity $m-1$, it does not affect the uniform or 1-up fixed points, hence the hypothesis $m > 1$.

We look for the special value of q such that $\lambda_2 = 0$ by solving a pair of quadratic equations: The first (24a) is (10) with Z replaced by X, which constrains X to be either $X^{(m)}$ or $Y^{(n-m)}$. The second (24b) is an expansion of $\lambda_2 = 0$ assuming $X = X^{(m)}$, that is, that we are evaluating λ_2 at an m-up fixed point. When fully expanded, these two quadratic equations are as follows:

$$\begin{aligned}&(a-1)m(n-1)X^2\\&+(a-1)(1+2m(q-1)-nq)X\\&-(a(1-m-n)-1)(q-1)=0,\end{aligned} \tag{24a}$$

$$\begin{aligned}&\left(-\frac{(1-a)mn}{n-m}\right)X^2\\&+\left(\frac{2(1-a)n(-1+m-mq+nq)}{(n-m)(n-1)}\right)X\\&+\left(-\frac{1-a}{n-m}-\frac{an(1-q)}{n-1}\right)=0.\end{aligned} \tag{24b}$$

It should be noted that there are solutions to this system that do not correspond to sign changes of λ_2 or to bifurcations in the symmetric language equation. These extraneous solutions will be eliminated once all solutions are found. Although one could conceivably substitute the explicit expressions for $X^{(m)}$ and $Y^{(m)}$ into the equation $\lambda_2 = 0$ hoping to solve it for q, the resulting equation has several embedded square roots, and in manipulating it to get rid of them, extraneous solutions are bound to appear. By dealing with this system instead, it is easier to prove certain things about the solutions that will ensure that we find all of them, and that we can determine which ones are extraneous.

The first result is that for each value of q which solves the system in question, there are at most two values of X, and for each value of X, there is at most one value of q. This is evident because when q is fixed, the two quadratic equations can have at most two common roots, and when X is fixed, both equations are linear in q.

The second result is that there are two possible values of q, which are found as follows. Multiplying (24a) by $-n$ and (24b) by $(n-1)(n-m)$ and adding the two results together yields, after much simplification:

$$-(1-a)(nq-1)(nX-1) = 0. \tag{25}$$

At this point, we have two choices, either $q = 1/n$, or $X = 1/n$. In the first case, we get two solutions for X because when $q = 1/n$ both quadratic equations turn out to have the same two roots; however, they are both complex, and are of no further interest. In the second case, the two quadratic equations in X become linear in q upon substituting $X = 1/n$, and we get a single solution

$$q = \frac{1 + n + a\left(n^2 - n - 1\right)}{n(2 + a(n-2))} = \hat{q}_{max}. \tag{26}$$

This is the unique parameter value for which λ_2 changes signs. To determine the signs, we plug the extreme case $q = 1, X = \frac{1}{m}$ into λ_2, which yields

$$\lambda_2\big|_{q=1, X=\frac{1}{m}} = \frac{1-a}{m},$$

which is positive. Therefore, λ_2 is negative for $q < \hat{q}_{max}$ and positive for $q > \hat{q}_{max}$.

It is important to notice that if $m < n/2$, then $X^{(m)} > 1/n$, so for these m-up fixed points, λ_2 is positive for all q such that the fixed points exist, and never changes sign. To prove this inequality, observe from Equations (11) and (13) that

$$X^{(m)} \geq \frac{1-q}{n-1} + \frac{nq-1}{2m(n-1)}.$$

For any fixed q, the term $(nq-1)/(2m(n-1)$ is minimized by making m as large as possible. If we require $m < n/2$, then

$$X^{(m)} > \left.\frac{1-q}{n-1} + \frac{nq-1}{2m(n-1)}\right|_{m=\frac{n}{2}} = \frac{1}{n}.$$

On the other hand, for $m > n/2$, the m-up fixed points always satisfy $X^{(m)} = 1/n$ at $q = \hat{q}_{max}$. To prove this, recall that (10) is a quadratic equation whose roots $X^{(m)}$ and $Y^{(n-m)}$ are numbers which appear m times as entries of m-up fixed points. It can be seen by substitution that if $q =$

$\hat{q}_{max}$, then $1/n$ is a root of this quadratic, so either $X^{(m)}$ or $Y^{(n-m)}$ has to be $1/n$. Assume that $m > n/2$ and $Y^{(n-m)} = 1/n$. It follows that $X^{(n-m)} = 1/n$ which yields a contradiction because $n-m < n/2$ and from a preceding argument $X^{(n-m)} > 1/n$. Therefore, $X^{(m)} = 1/n$ and $Y^{(n-m)}$ is the other root.

In the case where n is even and $m = n/2$, the m-up fixed points come into existence at $q = \hat{q}_m = \hat{q}_{max}$ and $X^{(m)} = 1/n$, so for them, $\lambda_2 = 0$ at that point and $\lambda_2 > 0$ for all larger q.

In summary, the sign change in λ_2 takes place for the m-up fixed points where $m > n/2$ and for no others. □

Proposition 4. *For an m-up fixed point where $n-1 > m > n/2$, the eigenvalue λ_3 is strictly positive for $q < \hat{q}_{max}$, zero for $q = \hat{q}_{max}$, and strictly negative for $q > \hat{q}_{max}$. If $0 < m \leq n/2$, then the eigenvalue $\lambda_3 \leq 0$ for $q = \hat{q}_m$ and strictly negative for $q > \hat{q}_m$.*

Proof. Since λ_3 is of multiplicity $n-1-m$, it does not affect $(n-1)$-up fixed points, hence the assumption that $n-1 > m$.

The analysis for λ_3 is quite similar to that for λ_2, and λ_3 is zero exactly when $q = \hat{q}_{max}$ and $X = 1/n$. It turns out that

$$\lambda_3\big|_{q=1, X=\frac{1}{m}} = -\frac{1-a}{m},$$

so λ_3 is positive for $q < \hat{q}_{max}$ and negative for $q > \hat{q}_{max}$. Again, the sign change in λ_3 takes place for the m-up fixed points where $m > n/2$ and for no others. □

Proposition 5. *For an m-up fixed point where $n > m > n/2$, the eigenvalue λ_1 is strictly positive for $\hat{q}_m < q < \hat{q}_{max}$, zero for $q = \hat{q}_m$ or $\hat{q}_{max}$, and strictly negative for $q > \hat{q}_{max}$. If $0 < m \leq n/2$, then the eigenvalue λ_1 is zero for $q = \hat{q}_m$ and strictly negative for $q > \hat{q}_m$.*

Proof. The analysis for λ_1 is also similar, but yields two sign changes. The two quadratic equations are

$$\begin{aligned}&(a-1)m(n-1)X^2\\&+(a-1)(1+2m(q-1)-nq)X\\&-(a(1-m-n)-1)(q-1)=0,\end{aligned} \tag{27a}$$

$$\begin{aligned}&\left(\frac{-3(1-a)mn}{n-m}\right)X^2\\&+\left(-\frac{2(1-a)(m+n-3mn+2mnq-n^2q)}{(n-m)(n-1)}\right)X\\&+\left(\frac{1-n-2m(1-q)-a(1-n+n^2-m(n+2)(1-q)-n^2q)}{(n-m)(n-1)}\right)=0.\end{aligned} \tag{27b}$$

The first one, (27a), is the same as (24a) and constrains X to be either $X^{(m)}$ or $Y^{(n-m)}$. The second one, (27b), is an expanded form of $\lambda_1 = 0$. The linear combination of $-3n$ times (27a) plus $(n-1)(m-n)$ times (27b) yields a large linear equation in X, which allows us to eliminate X in the first quadratic equation and find two values of q. The first turns out to be $q = \hat{q}_m$, which requires $X = X^{(m)}$ or $Y^{(n-m)}$. This is the bifurcation in which the m-up and $(n-m)$-up fixed points come into existence. The second is $q = \hat{q}_{max}$, which requires $X = 1/n$. Once again, this second sign change takes place for the m-up fixed points where $m > n/2$ and for no others. $\square$

For the specific case of $m = 1$, the only eigenvalues are λ_1 and λ_3. Only Propositions 4 and 5 are relevant, and prove that the 1-up fixed points are sinks when they exist. These three propositions together also prove that all other m-up fixed points are saddles of some kind.

7.2. Bifurcations of the uniform fixed point

The uniform fixed point, which is best thought of as the case where $m = 0$, is a special case, as it has only two distinct eigenvalues: λ_0, which is not of interest, and $\lambda_3 = c_1 - c_3$, which determines the stability of the fixed point. Again, we look for the special value of q that makes $\lambda_3 = 0$. The expression $c_1 - c_3 = 0$ evaluated at $x_j = 1/n$ yields a linear equation in q whose solution is the familiar

$$q = \frac{1 + n + a\left(n^2 - n - 1\right)}{n(2 + a(n-2))} = \hat{q}_{max}. \tag{28}$$

For $q < \hat{q}_{max}$, the uniform fixed point will be a stable sink, and for q any larger, it will be an unstable source.

7.3. Remarks about the bifurcations

Note that due to the symmetry of this dynamical system, $\hat{q}_{max}$ appears as a bifurcation point for many of the fixed points. As q increases to $\hat{q}_{max}$, all the fixed points come into existence, and for even n, the $n/2$-up fixed points come into existence right when $q = \hat{q}_{max}$. At this value of q, the m-up fixed points for $m > n/2$ all collide with the uniform fixed point in the center of the simplex. As q increases further, the fixed points all separate, with none being lost, but the uniform fixed point has completely reversed its stability. This behavior is known as an *$\mathcal{S}_n$-symmetric transcritical bifurcation*. (See [1].)

8. Other properties of the vector field

The vector field given by (8) can be written as the gradient of a function $V(x)$ plus an additional term. A number of well-known proofs [10,11] about gradient dynamical systems can be adapted to work on this ODE because of its near-gradient form, and the fact that the trajectories of interest are confined to a simplex.

Proposition 6. *Define the function $V(x)$ as follows:*

$$V = \frac{1}{3}(1-a)(q-u)M_3 - \frac{1}{4}(1-a)M_2^2 - \frac{1}{2}aunM_2 + auM_1. \tag{29}$$

If $x(t)$ is a trajectory of (8) which is confined to the simplex S_n and is not a fixed point, then the function $V(x(t))$ is strictly increasing as time advances.

Proof. This function was selected so that $\partial V/\partial x_j$ accounts for as many terms of the right-hand side of (8) as possible. Thus, (8) may be re-written as

$$\dot{x}_j = \frac{\partial V}{\partial x_j} + (1-a)uM_2,$$

or in vector notation:

$$\dot{x} = DV + (1-a)uM_2\mathbf{1}. \tag{30}$$

Computing the time derivative of V and using (30) to substitute for DV yields

$$\dot{V} = DV \cdot \dot{x} = \|\dot{x}\|^2 - (1-a)uM_2\mathbf{1} \cdot \dot{x}.$$

Since the trajectories of interest lie in the simplex, $\mathbf{1} \cdot \dot{x} = 0$, so the second term vanishes, leaving

$$\dot{V} = \|\dot{x}\|^2. \tag{31}$$

On any trajectory other than a fixed point, $\dot{x}$ will be non-zero, so $\dot{V}$ will be strictly positive. Therefore, V will be strictly increasing with time. □

This proposition implies the following:

Proposition 7. *The ODE given by (8) has no solutions which are periodic closed orbits, homoclinic loops, or directed heteroclinic cycles.*

Proof. Suppose $x(t)$ is a periodic closed orbit of period T, where $T > 0$. Then $x(0) = x(T)$, which implies

$$0 = V(x(T)) - V(x(0)) = \int_0^T \dot{V}\,dt.$$

However, the integrand is strictly positive, so the right-hand expression cannot be zero, and we have a contradiction.

A similar argument handles the cases of homoclinic loops and directed heteroclinic cycles as follows. Suppose $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m$, where $m \geq 1$, are fixed points, each of which is connected to the next by an orbit, and $\bar{x}_m$ is connected back to $\bar{x}_1$. By a similar argument, $V(\bar{x}_1) < V(\bar{x}_2) < \cdots < V(\bar{x}_m) < V(\bar{x}_1)$, so we have a contradiction. □

Proposition 8. *All orbits of (8) tend to some fixed point as $t \to \infty$.*

Proof. The function $V(x)$ is continuous and its domain, the simplex, is compact. Therefore, $V(x)$ is bounded. For an orbit $x(t)$ other than a fixed point, the value of $V(x(t))$ is strictly increasing and bounded. It must therefore approach a finite limit from below as $t \to \infty$. This implies that $\dot{V} \to 0$, and by (31), $\dot{x} \to 0$, which is only possible if the orbit converges to a fixed point. □

9. Conclusion

The results of the preceding sections allow us to form a fairly complete picture of the fully symmetric language dynamical equation. In particular, we have a complete description of the pattern of bifurcations as q increases from $1/n$ to 1. For low values of q, there is only one fixed point, the uniform fixed point, and it is a stable sink. As q exceeds $\hat{q}_1$, the 1-up and $(n-1)$-up fixed points appear in pairs, one pair in each corner of the simplex, through saddle-node bifurcations. The 1-up fixed points are stable sinks and remain stable as q increases, but the $(n-1)$-up fixed points are saddles. Their stable manifolds initially form the boundaries between the basins of attraction of the uniform fixed point and those of the 1-up fixed points. As q increases further, the other m-up fixed points appear in saddle-node bifurcations, and are always saddles. When q finally reaches $\hat{q}_{max}$, the m-up fixed points for $m > n/2$ all collide with the uniform fixed point in an $\mathcal{S}_n$-symmetric transcritical bifurcation. As q increases, the m-up fixed points separate, having shuffled their stabilities and become saddles with different stable and unstable manifolds than they had before the bifurcation. The uniform fixed point continues to exist, but is now an unstable source.

By analyzing the fully symmetric language dynamical equation as a near-gradient system, we have shown that its behavior is fairly straightforward. There are no closed orbits, no homoclinic loops or directed heteroclinic cycles, and all orbits tend to a fixed point as time increases.

This model provides a mathematical foundation for understanding linguistic phenomena that have to do with population dynamics, evolutionary phenomena, and learning in heterogeneous environments. Consider for example the transition from Old English to Middle English. One theory, described in [16], is that part of the change was due to the influence of Scandinavian invaders, whose language was somewhat similar to Old English. The results in this paper may be linked to this hypothesis as follows. Before the invaders, the English linguistic environment was relatively uniform apart from dialectical differences. In this situation, the grammars available to children were a number variations of Old English, and learning was very reliable, that is, a was close to 1 but q was large enough that the population had settled into a single-grammar equilibrium. (Grammars very different from Old English can be ignored, as the probability that a child makes enough learning errors to speak something totally different seems to be tiny in this case.) Once the invaders arrived, the presence of Scandinavian speech caused enough confusion that children were unable to properly acquire certain features of Old English, such as the case system. That is, the added linguistic noise caused q to decrease enough to destabilize the single-grammar equilibrium. When the invasion ceased, q increased again, and the population settled down into a different single-grammar equilibrium.

The results of this paper can be extended in a number of directions. For example, when the A and Q parameters of the language dynamical equation are set to the symmetric forms here plus a small, asymmetric perturbation, the mass collision of fixed points which results in the $\mathcal{S}_n$ transcritical bifurcation will not occur, and the transition to coherence will happen in several small bifurcations instead of one big one. To understand all possible perturbations would require finding a universal unfolding of the $\mathcal{S}_n$ transcritical bifurcation in an arbitrary number of dimensions. (See, for example, [1] for another instance of this bifurcation and [7] for a relevant theorem.) An alternative would be to explore parameter settings which have a smaller symmetry group, for example, cyclic or dihedral symmetry. This model currently ignores spatial effects, such as clustering, which are known to be crucial in maintaining linguistic diversity. For example, Papua New Guinea is home to many isolated tribes separated by mountains, and hundreds of languages are spoken there. Spatial effects could be incorporated into this model by re-formulating it for a number of discrete cities with limited interaction, or by moving to a system of partial differential equations.

References

1. P. Ashwin and J. W. Swift. The dynamics of n weakly coupled identical oscillators. *Journal of Nonlinear Science*, 2:69–108, 1992.
2. D. Bickerton. *Roots of Language*. Karoma Publishers, Inc., Ann Arbor, 1981.
3. N. Chomsky. *Language and Mind*. Harcourt Brace Jovanovich, New York, 1972.
4. N. Chomsky. *Language and Problems of Knowledge*. MIT Press, 1988.
5. M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, 1971.
6. M. Eigen, J. McCaskill, and P. Schuster. The molecular quasi-species. In I. Prigogine and S. A. Rice, editors, *Advances in Chemical Physics*, volume 75, pages 149–263. John Wiley and Sons, New York, 1989.
7. M. J. Field and R. W. Richardson. Symmetry breaking and the maximal isotropy subgroup conjecture for reflection groups. *Archive for Rational Mechanics and Analysis*, 105:61–94, 1989.
8. E. Gibson and K. Wexler. Triggers. *Linguistic Inquiry*, 25:407–454, 1994.
9. E. M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.
10. J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer-Verlag, 1990.
11. M. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, 1997.
12. J. Hofbauer and K. Sigmund. *Evolutionary Games and Replicator Dynamics*. Cambridge University Press, 1998.
13. N. L. Komarova, P. Niyogi, and M. A. Nowak. The evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology*, 209(1), 2001.
14. N. L. Komarova and M. A. Nowak. The evolutionary dynamics of the lexical matrix. *Bulletin of Mathematical Biology*, 63(3):451–485, 2001.
15. D. Lightfoot. *How to Set Parameters: Arguments from Language Change*. MIT Press, Cambridge, MA, 1991.
16. D. Lightfoot. *The Development of language: Acquisition, Changes and Evolution*. Blackwell Publishers, 1999.
17. M. A. Nowak and D. C. Krakauer. The evolution of language. *Proceedings of the National Academy of Science, USA*, 96:8028–8033, 1999.
18. M. A. Nowak, J. Plotkin, and V. A. A. Jansen. Evolution of syntactic communication. *Nature*, 404:495–498, 2000.

19. M. A. Nowak, J. Plotkin, and D. C. Krakauer. The evolutionary language game. *Journal of Theoretical Biology*, 200:147–162, 1999.
20. S. Pinker. *The Language Instinct*. W. Morrow and Company, New York, 1990.
21. A. Senghas and M. Coppola. Children creating language: How Nicaraguan sign language acquired a spatial grammar. *Psychological Science*, 12(4):323–328, July 2001.