A Mathematical Model of Human Languages: The Interaction of Game Dynamics and Learning Processes

William Garrett Mitchener

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Program in
Applied and Computational Mathematics

June 2003

Abstract

Human language is a remarkable communication system, apparently unique among animals. All humans have a built-in learning mechanism known as *universal grammar* or UG. Languages change in regular yet unpredictable ways due to many factors, including properties of UG and contact with other languages. This dissertation extends the standard replicator equation used in evolutionary biology to include a learning process. The resulting language dynamical equation models language change at the population level. In a further extension, members of the population may have different UGs. It models evolution of the language faculty itself.

We begin by examining the language dynamical equation in the case where the parameters are fully symmetric. When learning is very error prone, the population always settles at an equilibrium where all grammars are present. For more accurate learning, coherent equilibria appear, where one grammar dominates the population. We identify all bifurcations that take place as learning accuracy increases. This alternation between incoherence and coherence provides a mechanism for understanding how language contact can trigger change.

We then relax the symmetry assumptions, and demonstrate that the language dynamical equation can exhibit oscillations and chaos. Such behavior is consistent with the regular, spontaneous, and unpredictable changes observed in actual languages, and with the sensitivity exhibited by changes triggered by language contact.

From there, we move to the extended model with multiple UGs. The first stage of analysis focuses on UGs that admit only a single grammar. These are stable, immune to invasion by other UGs with imperfect learning. They can invade a population that uses a similar grammar with a multi-grammar UG. This analysis suggests that in the distant past, human UG may have admitted more languages than it currently does, and that over time variants with more built-in information have taken over.

Finally, we address a low-dimensional case of competition between two UGs, and find conditions where they are stable against one another, and where they can coexist. These results imply that evolution of UG must have been incremental, and that similar variants may coexist.

This research was conducted under the supervision of Dr. Martin A. Nowak (Program in Theoretical Biology at the Institute for Advanced Study, and Program in Applied and Computational Mathematics at Princeton University).

Acknowledgments

Thanks to my advisor, Dr. Martin Nowak, for guiding me to such an interesting area of research. Thanks to Dr. Marguerite Browning, and the other members of the Program in Linguistics for their support. Thanks to Dr. Burton Singer for reading my dissertation, and to Dr. Philip Holmes and Dr. Simon Levin for being on my examination committee. And thanks to the Program in Applied and Computational Mathematics for finding support for my graduate studies.

Thanks to Dr. Natalia Komarova and all my other friends and co-workers from the Program in Theoretical Biology at the Institute for Advanced Study.

I must also thank Cynthia Rudin, Eric Brown, Dr. Jeffrey Moehlis, and all my other friends from PACM for their support and encouragement.

Special thanks to my long-time friend and mentor Dr. Harold Reiter.

Last but not least, I wish to thank Mom and Dad and Allison and Ginger for all their love and support.

Contents

Chapter 1

Introduction and Background

Contents

This dissertation is the product of several years of interdisciplinary research, combining ideas from theoretical biology, linguistics, computer science, and psychology all under the umbrella of applied mathematics. One of the pleasures of interdisciplinary research is the opportunity to learn exciting things about other fields, linguistics and biology in my case. My initial impression of linguistics from several years ago was that it must be a horribly dull subject, associated with images of well-meaning teachers drilling the “proper” ways of speaking and writing into defenseless students who cannot overcome their urges to split infinitives and end sentences with prepositions. Needless to say, I quickly changed my mind, and linguistics has turned out to be one of my favorite subjects, full of experts who in fact take delight in splitting infinitives and ending sentences with prepositions while attempting to understand why such constructions are perfectly grammatical. Likewise, mathematical biology is a very rich field, growing out of the need for a more precise understanding of increasingly complex observations and experimental data.

I will begin with a section introducing linguistics, focused on the subject of language change. The next section gives background information about replicator dynamics, which is the mathematical framework that will be extended in the rest of this dissertation. I will conclude by summarizing my results, and connecting the background information in this introduction to the material to follow.

1.1. Linguistics and language change

Human language is a remarkable communication system, apparently unique among animals in its overall complexity [27, 28]. Bird songs, for example, are thought to all mean the same thing despite their variation: "This is my spot and I'm available." Vervet monkeys have fixed alarm calls for various predators, but these seem to be innate rather than learned and lack the flexibility of human vocabulary. Chimpanzees and other apes can learn to communicate through simplified forms of sign language, but do not in general make much use of syntax, such as word order and relative clauses, as is central to human language.

Human language may be roughly divided into two parts, although there is considerable debate over exactly what goes in which part [34]. The *lexicon* is a table of words, parts of words, and phrases, together with their associated meanings. The *grammar* is the set of rules for combining lexical items into more complex phrases and sentences to express composite meanings.

1.1.1. Parts of grammar. Grammar consists of several sub-divisions, and again there is considerable debate over exactly what goes into each. For the purposes of this dissertation, it suffices to divide grammar into the areas of phonology, morphology, syntax, and semantics.

The *phonology* of a language is the set of rules of allowed sound combinations, syllable construction, and stress placement.[1] Each language has a set of rules for determining what combinations of sounds are acceptable. For example, *plooper, stad,* and *blicket* are distinctly English nonsense words, conforming to the rules of English phonology, and *dongle* and *spam* are distinctly English words that have taken form within only the last century. On the other hand, the names *Svetlana, Dvořák,* and *Nguyen* are distinctly non-English because they violate English phonology while conforming to the phonology of their original language (Russian, Czech, and Vietnamese, respectively): English does not make use of the consonant mixture /sv/, the trill[2] /ř/, or syllables beginning with the consonant /ŋ/.

The *morphology* of a language is the set of rules for forming words out of fragments called *morphemes.* For example, the prefix *re-* on a verb changes its meaning to include the idea of "doing again" or "putting back to an earlier state," as in *repaint* or *replace.* English also allows the formation of an endless variety of compound words: *underwear,*

[1] A large set of symbols known as the *International Phonetic Alphabet* or *IPA* has been developed for precisely recording pronunciations without the confusion of language-specific alphabets and spelling conventions. Characters in phonemic slashes, as in /tɪp/ for the word *tip,* represent fairly loose transcriptions capturing only the most important information. More detailed transcriptions are written in phonetic braces, as in [ˈtʰɪp].

[2] According to Dr. Mirjam Fried of Princeton University, this is an accepted phonetic symbol for this sound.

earring, bellybutton, Christmas tree, guard dog, joystick, speed bump, pocketbook, good-for-nothing. French morphology is less accepting of pure compound words, preferring to include prepositions and form a short phrase instead: *soupe à l'oignon, coq au vin, sac à main, bon à rien.*

The *syntax* of a language is the set of rules for forming phrases and sentences. In English, the subject typically comes first, followed by the verb and its complement. In Japanese on the other hand, verbs follow their complements. French makes use of small words called *special clitics* for pronouns that are objects of verbs, and they must precede the verb in a specific order. Syntax also include rules for determining binding of pronouns such as *him* and anaphors such as *himself* to antecedents:[3]

(1.1.1) John_1 likes pictures of himself_1. (where *himself* refers to *John*)

(1.1.2) *Mary likes pictures of himself. (where *himself* refers to a man in the surrounding discourse)

Many languages include syntactic rules for moving words around, as in the formation of questions from statements:

(1.1.3) Who_1 did you see t_1?

(1.1.4) *Who_1 do you wonder whether John saw t_1?

(1.1.5) Do you know the muffin man?

(1.1.6) *Know_1 you t_1 the muffin man? (acceptable in 17th century forms of Modern English)

The t_1 is called a *trace* and is an unpronounced place holder left behind when something moves.

Semantics is the interface between the raw structure of language and meaning. For example, the following sentences are grammatical:

(1.1.7) The man climbs.

(1.1.8) The man rises.

but there are significant semantic differences between *climb* and *rise*, as illustrated by:

(1.1.9) The man climbs down.

(1.1.10) *The man rises down.

The study of semantics very quickly becomes intertwined with theories of knowledge and philosophy [34]. Such tangled subjects will be avoided for the most part in this dissertation as they are not immediately relevant.

The idea of a language also spans multiple scales. On the large end, the term *English* describes the native speech patterns of most people in the United Kingdom, the United States, Canada, Australia, New Zealand, and India, spanning the past several centuries.

[3] In linguistic notation, a * indicates an ungrammatical or otherwise ill-formed example. A ? indicates a grammatically not-quite-right example. Subscripted indices are used to indicate that two elements of a sentence are co-referential.

Within the world-wide English-speaking community there are any number of regional, social, and age-group dialects, all with slightly different grammars. At the finest level, each individual may use the language a little bit differently at different times and in varying situations. The analogy with biology is that *language* is something like *species*, in that both terms describe a group of non-identical individuals with much in common.

1.1.2. Acquisition. Children acquire their native language by hearing example sentences from their parents through which they learn both the lexicon and the grammar [12, 45, 61]. Grammar acquisition is thought to be based solely on positive evidence, in that children attempt to build a grammar that is somehow consistent with what they hear, while ignoring explicit instructions and corrections from adults [44]. Negative evidence, meaning information that a particular sentence is not grammatical, is discarded. The general problem of acquiring grammar only from example sentences is known to be impossible without constraints on the rules of the grammar [24]. In short, any finite set of sentences is insufficient to specify a unique grammar. For example, the sample set $S = \{s_1, s_2, \dots, s_n\}$ of finite strings of symbols from a lexicon Σ is consistent with the grammar containing just the sentences of S and with the grammar containing all finite strings of symbols from Σ. Hence, linguists hypothesize that the human mind includes a set of innate constraints and hints known as *universal grammar* or *UG* that provide additional information to children and guide them to acquire one particular grammar [11, 61].

UG operates even when the input is exceptionally impoverished, as in the cases of creolization [7] and the spontaneous invention of sign languages [68]. In both of these cases, the example input comes from a more or less *ad hoc* and artificial system of communication that does not conform to UG. 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 requirement that nouns be assigned case, appear to be universal across all languages. Other 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 [12], and learning these parameter settings is equivalent to choosing among a finite number of possible classes of grammars [23]. This model is known as the *principles and parameters* framework. Phonological rules are better modeled by systems of constraints that may override one another, and learning such a system involves determining the priorities of the different constraints. This model is known as *optimality theory* [72].

1.1.3. Language change. Children do not acquire their native language perfectly. Given that languages include variation in time, situation, and speaker, it is not even clear what perfect learning would mean. Instead, children may make certain “mistakes” in acquiring the language of the preceding generation. These mistakes can be caused by any number of factors, such as style, systematic phonological changes, or contact with another language. Over time, small changes can accumulate and transform a language significantly. Every so often, a catastrophic change occurs.

I will begin by describing some examples of fairly recent changes in English, based on common knowledge and some personal observations. From there I will proceed to some examples of older, more drastic, and more carefully studied changes in English over the past several centuries. Finally I will discuss some changes in other languages. These examples illustrate that language change may be regular while remaining unpredictable, and how seemingly minor changes to the linguistic environment can lead to large-scale language change.

The pronoun *whom* is the equivalent of *who* but with objective (or accusative) case:

(1.1.11) Whom did you see?

(1.1.12) Who saw you?

But after many generations, it has simply fallen out of use, and now, *who* is typically used in both cases, unless the speaker is making an effort to be formal:

(1.1.13) Who did you see?

Another example concerns the influence of style on grammar. Most school children are successfully taught that when mentioning oneself as part of the subject of a sentence, it is proper to put the *I* last:

(1.1.14) My friend and I went to the store.

(1.1.15) ?I and my friend went to the store.

People often extend this rule to objects as well. Instead of unquestionably correct sentences using the object pronoun *me*:

(1.1.16) Alice talked to me and Jim.

(1.1.17) Alice talked to Jim and me.

it is common to use the subject pronoun *I* instead:[4]

(1.1.18) †Alice talked to Jim and I.

English grammar does not seem to require nouns joined by *and* to have any particular case, and although *me* is the obvious choice for an object, many people choose *I,* perhaps because of style [61, p. 390]. Another example from Modern English appears to be due to the frustration that in some dialects (those lacking *ain't*) there is no contraction for *am not.* Hence the following:

(1.1.19) I'm right, am I not?

(1.1.20) †I'm right, aren't I?

(1.1.21) ?Aren't I right?

(1.1.22) *Are I right?

(1.1.23) *I aren't right.

(1.1.24) *I are right.

This seems to be possible because the verb *be* is completely irregular. Every form has slightly different properties, and seems to be stored as a separate item in the lexicon [45, § 7.1]. A transformation from style to grammar centuries ago may also be the origin of

[4] I will use † to indicate an example that seems to be grammatical according to common usage despite some apparent problem.

do as an auxiliary verb, replacing *Thou sing'st* with *Thou dost sing*, which was easier to pronounce and more conducive to poetry [45, p. 97].

Let us continue with some older changes in English. The history of English has been traced back roughly five or six thousand years, all the way to Proto-Indo-European (or PIE), the ancestor of most languages spoken in Europe, the Mediterranean area, the Middle East, and India. Over time, changes took place in one location but not another, and the result was vast diversification and the present variety of Indo-European languages. The vocabulary of PIE has been reconstructed by comparing words from a wide range of languages and guessing the most likely phonological changes that lead from one to the other.

In parts of Europe, PIE developed into Old Germanic, which later split into North, East, and West Germanic. English, Dutch, and German are descendants of West Germanic [13, 74, 80]. Or rather, this is roughly what seems to have happened. The actual history of Germanic languages seems to be more complicated, with some parallel changes happening in different languages and words being borrowed, so the lineage is not completely clear [79]. The transition from PIE to Old Germanic included several regular phonological changes known as Grimm's Law and Verner's Law. Grimm's Law says that an unvoiced stop in PIE becomes the corresponding fricative in Old Germanic, so for instance /p/ becomes /f/. Furthermore, a voiced stop becomes the corresponding unvoiced stop, and an aspirated voiced stop becomes unaspirated. Verner's Law, developed to explain what were originally thought to be exceptions to Grimm's Law, says that unvoiced stops become voiced stops when preceded by an unstressed syllable. The final exception is that no change occurs to consonants preceded by an unvoiced consonant. So for example, the PIE word[5] /pə$_2$ˈter/ became the Old English word /ˈfæder/, and eventually the Modern English /ˈfaðər/ or *father* [80]. The original PIE word was transferred to Greek as /paˈter/ with the consonants and stress intact. Changes such as these have been noted in any number of the world's languages. Although phonological changes are often very regular, it is not clear why they happen, or when they will happen. So despite its regularity, this form of language change is still unpredictable.

The English case system used to be as rich as that of any other Germanic language. In fact, Old English used to inflect its equivalent of *that stone* as follows:[6]

	Singular	Plural
Nominative	se stān	þā stānas
Accusative	þone stān	þā stānas
Genitive	þæs stānes	þāra stāna
Dative	þæm stāne	þæm stānum

Old English was in use until about 1066 A.D. (the invasion of England by the Normans). Around that time, the language transformed into Middle English, and part of the transformation was the loss of this rich case system. One theory is that the loss was caused (or perhaps accelerated) by the presence of Scandinavian invaders who spoke Old Norse, a

[5] The /ə$_2$/, also written /h$_2$/, is one of the disputed laryngeal sounds in PIE.

[6] The Old English character þ named *thorn* represents the sound [θ] of an unvoiced *th* as in *thick*, except when surrounded by voiced sounds where it represents the sound [ð] of a voiced *th* as in *them*. The symbol æ has the sound /æ/ of the vowel in *at*. The macron ¯ over a vowel means it is pronounced for a longer duration.

similar Germanic language with a similar case system. The theory says that children heard both sets of case endings and were unable to acquire either one, so they started leaving out the case endings and the system disappeared.

This example brings up an interesting point. One of the questions that tends to come up concerning this research has to do with bilingualism: Children brought up in a multilingual environment appear to have no trouble acquiring several native languages. It is therefore something of a paradox that language contact should have such consequences. The paradox may be due to some kind of phase transition. If children are exposed to several clearly distinguishable languages, they can acquire them all. If children are only exposed to variations of one language, they overlook those variations and acquire some average form of the language as spoken by the whole population. However, the speech of two individuals may be too different to be considered variations of one language, but too similar to be considered separate languages. If children are brought up in such an environment, it is not clear what language they will acquire, as there is no strong push toward either bilingualism or monolingualism. It is therefore possible that they will end up with some language that is different from either of the ones in the environment.

As a last example, there are roughly three types of morphology systems in the world's languages. An *isolating* language has many small words that each carry one small fragment of meaning, and sentences are formed by combining many small words. Vietnamese is just such a language:[7]

(1.1.25) Khi tôi dến nhà bạn tôi, chúng tôi bắt dầu làm bài
when I come house friend I, PLURAL I begin do lesson
"When I arrived at my friend's house, we began to do lessons." [74, p. 126]

Observe that *we* is expressed by a plural word plus the equivalent of *I*. In an *agglutinating* language, words are formed from a stem plus many small morphemes, each of which carries a fragment of meaning, as in Turkish:

(1.1.26) Yap-tiğ-ım hata-yı memleket-i tanı-ma-ma-m-a ver-ebil-ir-siniz
make-PART-my mistake-OBJ country-OBJ know-not-GER-my-to give-can-TENSE-you
"You can ascribe the mistake I made to my not knowing the country." [74, p. 126]

The third morphology type is *inflecting,* in which a single morpheme carries several fragments of meaning. The typical example is Latin:

(1.1.27) Arm-a vir-um-que can-ō.
weapon-NEUTPLURALOBJ man-MASCSINGOBJ-and sing-FIRSTPRESINDICACT
"Arms and the man I sing." [74, p. 126]

An isolating language can change into an agglutinating language by running all those short words together. An agglutinating language can become inflecting through phonological changes in which commonly combined morphemes are contracted into a new morpheme

[7] The abbreviations used here for syntactic information are as follows: SING–singular, PLURAL–plural, MASC–masculine, NEUT–neuter, FIRST–first person, OBJ–object, PART–participle, GER–gerund, TENSE–tense, PRES–present tense, INDIC–indicative mood, ACT–active.

with a compound meaning. Inflecting languages can become isolating through a process where the initial or final sounds of words are systematically reduced, thereby erasing morphemes. The result is a cycle between the three morphology types. The changes in this cycle are generally so slow, requiring perhaps thousands of years, that no complete cycle has been observed, however, instances of languages changing from one type to another have been found. It certainly seems possible for languages to make the reverse transformations, but the general trend seems to be isolating to agglutinating to inflecting and back to isolating. Furthermore, many languages such as Navaho have maintained their morphology type despite other changes for a very long time. Thus, the morphology type cycle is a fairly regular language change that is still unpredictable.

1.1.4. Language and evolution. Language changes on geological as well as historical time scales, in that the language faculty within the brain itself changes through genetic mutation and natural selection. The study of the evolution of language has a messy history, an infamous episode of which came in 1866 when the Société de Linguistique in Paris banned papers on the subject. There is essentially no evidence for what human language was like more than a few thousand years ago. Reconstructions as for Proto-Indo-European are convincing, but fragmentary and limited to a few thousand years. Once we begin asking questions about the language capacities of hypothetical ancestor species, there is essentially no data other than anatomical indications from fossils that they might have been able to pronounce certain sounds. Hence, the subject is vulnerable to ungrounded speculation. Furthermore, the genetic encoding of the language ability is very poorly understood. It is not even known how much genetic variation there is in the human language faculty across the entire population. It may be that several significant variations exist among humans, or perhaps the need to communicate limits UG to inconsequential variations.

Recent advances in neurology, psychology, and genetics have sparked new interest in language evolution. For example, a gene has been linked to a specific language impairment [18, 42], confirming that there is at least some variation in UG. New ideas in linguistics, artificial intelligence, simulation, and mathematical modeling have led to an extensive literature and more precise theories [2, 8, 22, 25, 27, 29, 32, 33, 39, 41, 43, 61, 62, 65, 71].

1.1.5. Opportunities for mathematical modeling. Language change is a complex subject, involving interactions of effects at multiple scales in time and space, and mathematical models can be used to explore how these interactions work. There have already been several models of the lexicon [19, 37, 59, 73], the origins of grammar [54, 57, 58, 63], grammar acquisition [4, 5, 23, 24, 50, 51, 56, 72, 77, 78], and evolutionary dynamics of grammar [36, 38, 52, 55], as well as linguistic simulations [9, 10, 35]. There are any number of starting points from which to proceed in formulating such models. Models of the actual learning process operate at the scale of the individual. Alternatively, we could begin at the population level with a highly simplified model of learning, and ask questions about the overall behavior of the population. This dissertation takes the latter approach. In particular, we would like to address the following questions:

What causes a population to come to a consensus on a grammar? That is, given an initial state of a large population that includes speakers of many grammars, under what circumstances do they reach a state after several generations in which most individuals use a common grammar? What circumstances result in an equilibrium where multiple grammars remain present? This is a fundamental question for modeling the tendency of languages to remain largely unchanged for decades or centuries. This question will be addressed in Chapter 2 where it will be shown that single-grammar equilibria are possible only if learning is sufficiently accurate. This equilibrium result can still be used to understand language change: A language can be stable until some outside influence changes the conditions of the model to a state where a single grammar state is no longer stable. When the influence is removed, the population may once again settle on a single grammar, perhaps different from the original one.

Can a model based on the idea that survival depends on communication reproduce the phenomena observed in language change? By linking survival to communication, we can build on the existing frameworks of evolutionary game theory and population dynamics. This question will be addressed in Chapter 3. Note that the goal at this point is to reproduce certain phenomena, in particular, spontaneous fluctuations, systematic yet unpredictable changes, and sensitivity, as described in this section. Reproducing a particular change in detail would require a more elaborate model beyond the scope of this dissertation, as well as very thorough data that may not be available.

What can such a model tell about the evolution of the language faculty? Assuming that universal grammar has been shaped by natural selection, what can be determined about it? Chapter 4 investigates the question of why UG admits more than one grammar. Chapter 5 addresses more general questions of multiple UGs within a population, providing the start of a general framework for understanding when a new UG can invade an established UG, and when multiple UGs can coexist stably.

There are many other questions that could be addressed, concerning for example the rate of spread of language change [40], the phase boundary for a particular learning algorithm between those grammars that are too similar to be distinguished and those that are clearly different, spatial effects and the preservation of regional dialects, multilingualism, and so on. These issues may be addressed by future research.

1.2. Biological models of population game dynamics

In this section, I will give a brief review the replicator equation, a well-studied model of population dynamics where survival is based on an individual's expected payoff in an abstract game. To model interacting languages, we set up a communication game where grammars are the strategies. The resulting dynamical system has very restricted behavior, and cannot directly model language change. In Section 1.3, I will describe an extended model based on replicator dynamics that can model language change.

1.2.1. The replicator equation. The replicator equation is a thoroughly studied model of population dynamics under natural selection driven by an abstract game. Hofbauer and Sigmund [31] is a fairly complete reference, and we will begin with a summary of some important results from it.

Consider a large population, with n strategies available for use in an abstract game. Each individual must play one of these strategies. The fraction of the population using strategy i is denoted x_i, and we require $\sum x_i = 1$. Each fraction must satisfy $0 \le x_i \le 1$. Therefore, the population state may be represented as a point in the n-vertex simplex S_n, defined as follows:

$$S_n = \left\{ x \in \mathbf{R}^n : x_i \ge 0, \sum_{i=1}^{n} x_i = 1 \right\}. \tag{1.2.1}$$

The population changes according to a measure of fitness. The fitness of strategy i is given by F_i, which may take on a variety of forms. Hofbauer and Sigmund [31] considers cases where each F_i is a function of the population state x. The average fitness of the population is defined to be

$$\phi = \sum_{i=1}^{n} F_i x_i. \tag{1.2.2}$$

The rate of change of x_i with respect to time is modeled as being proportional to x_i and to how far the fitness F_i exceeds the average fitness, yielding the system of differential equations

$$\dot{x}_i = x_i(F_i - \phi) \text{ where } i = 1 \dots n. \tag{1.2.3}$$

Observe that if $x_i = 0$ at time 0, then it remains 0 for all time, so the sub-simplexes that form the boundary of the simplex are invariant. Furthermore, let $M_1 = \sum x_i$. Then

$$\begin{aligned} \dot{M}_1 &= \sum_{i=1}^{n} \dot{x}_i \\ &= \left(\sum_{i=1}^{n} x_i F_i \right) - \phi M_1 \\ &= \phi(1 - M_1). \end{aligned}$$

Thus, if the total population starts at $M_1 = 1$, it will remain so for all time, so the plane containing the simplex is invariant. All orbits of interest are therefore trapped inside the simplex.

It is typical to use a linear function of the x_j's for F_i:

$$F_i = \sum_{j=1}^{n} B_{i,j} x_j, \text{ or } F = Bx, \tag{1.2.4}$$

where B is interpreted as a payoff matrix: $B_{i,j}$ is the payoff to a player of strategy i against a player of strategy j. The fitness of strategy i in this case is an average of all possible payoffs to strategy i weighted by the appropriate fraction of the population. The average payoff simplifies to

$$\phi = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i B_{i,j} x_j = x^T B x.$$

The payoff matrix B represents the abstract game that drives the dynamics. There are several invariants: If a constant is added to each entry in a column of B, the overall behavior is preserved. Furthermore, if B is multiplied by a positive constant, the time can be rescaled to restore the original dynamics. It is helpful to re-write the differential equation in the following form:[8]

$$\dot{x}_i = x_i \left(e_i^T Bx - x^T Bx \right). \tag{1.2.5}$$

This dynamical system is equivalent to the Lotka-Volterra equation via a change of variables [31, sec. 7.5].

A point $x \in S_n$ represents a population state consisting of a particular mixture of the available strategies. A *Nash equilibrium* is a point $\hat{x} \in S_n$ such that

$$\forall x \in S_n : \hat{x}^T B\hat{x} \geq x^T B\hat{x}. \tag{1.2.6}$$

The term *Nash equilibrium* may also refer to a strategy that does at least as well against itself as every other strategy. Definition (1.2.6) is an extension of that idea to population states. Likewise, an *evolutionarily stable strategy* is one that cannot be invaded by any other strategy. This concept can be extended to also be extended to population states: An *evolutionarily stable state* or *ESS* is a point $\hat{x} \in S_n$ that satisfies

$$\hat{x}^T Bx > x^T Bx \tag{1.2.7}$$

for all $x \neq \hat{x}$ in some neighborhood of $\hat{x}$.

We will need the following results, which may be proved using the Lyapunov function

$$P(x) = \prod_{i=1}^{n} x_i^{\hat{x}_i}. \tag{1.2.8}$$

Theorem 1.2.1 (7.2.4 in [31])**.** *If $\hat{x} \in S_n$ is an ESS, then it is an asymptotically stable fixed point.*

Corollary 1.2.2 (p. 71 in [31])**.** *If $\hat{x} \in S_n^\circ$ is an ESS on the interior of S_n then it is a globally stable fixed point, meaning all orbits in S_n° converge to $\hat{x}$ in forward time.*

1.2.2. A communication game. For a communication game, the available strategies are the grammars $G_1, G_2, \ldots, G_n$. 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. The principles and parameters framework and optimality theory both imply a finite number of grammars. We assume further that each grammar communicates best with itself, and all payoffs are positive. The payoff matrix B is therefore diagonally dominant, meaning each diagonal element is greater than each other element in its row and column.

Replicator dynamics for such a game are fairly constrained. In particular, the vertices e_k of the simplex are stable fixed points. Substitution into (1.2.3) demonstrates at once that they are fixed points, and the following proposition shows that they are asymptotically stable.

[8] Recall that e_k represents a fundamental basis vector: All entries zero except the k-th, which is 1.

Proposition 1.2.3. *Suppose B is diagonally dominant. Then e_k is an ESS, and consequently, asymptotically stable.*

Proof. We must show that $e_k^T Bx - x^T Bx > 0$ for all $x \in S_n$ sufficiently close to e_k. First, observe that

$$e_k^T Bx - x^T Bx = \left(\sum_{j=1}^{n} B_{k,j} x_j \right) - \left(\sum_{i=1}^{n} \sum_{j=1}^{n} x_i B_{i,j} x_j \right).$$

The two terms can be combined if the first term is multiplied by $1 = \sum_i x_i$, which gives

$$e_k^T Bx - x^T Bx = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i (B_{k,j} - B_{i,j}) x_j.$$

All the terms with $i = k$ are zero, and we will need to treat the terms with $j = k$ separately, so the sum may be rewritten as

$$e_k^T Bx - x^T Bx = \left(x_k \sum_{i \neq k} (B_{k,k} - B_{i,k}) x_i \right) + \left(\sum_{i \neq k} \sum_{j \neq k} x_i (B_{k,j} - B_{i,j}) x_j \right).$$

To proceed, we need two new constants:

$$c = \min_i B_{k,k} - B_{i,k} > 0.$$
$$d = -\min_{i,j} B_{k,j} - B_{i,j}.$$

The diagonal dominance of B guarantees that $c > 0$, but d may be of either sign. Substituting c for $B_{k,k} - B_{i,k}$ in the first sum and $-d$ for $B_{k,j} - B_{i,j}$ yields the following inequality:

$$\begin{aligned} e_k^T Bx - x^T Bx &> x_k c \sum_{i \neq k} x_i - d \left(\sum_{i \neq k} x_i \right) \left(\sum_{j \neq k} x_j \right) \\ &= c x_k (1 - x_k) - d(1 - x_k)^2. \end{aligned}$$

If $d \leq 0$, then we are done, as the resulting expression would be positive. If not, then we must perform one last transformation to arrive at

$$e_k^T Bx - x^T Bx > (1 - x_k)((c + d)x_k - d),$$

which is positive for all x close enough to e_k that

$$x_k > \frac{d}{c + d}.$$

Either way, e_k is an ESS. Asymptotic stability follows immediately from Theorem 1.2.1. □

Corollary 1.2.4. *If B is diagonally dominant, the only evolutionarily stable states are the corners.*

Proof. Corollary 1.2.2 implies that there can be no ESS in S_n°, because trajectories near the corners converge to a corner and an interior ESS would be globally asymptotically stable.

Having an ESS on the boundary is ruled out by focusing on the sub-simplex in question, and observing that it is governed by lower-dimensional replicator dynamics under a sub-matrix of B. Since the sub-matrix inherits diagonal dominance, the above argument applies, and there can be no ESS. □

It is still possible to have asymptotically stable fixed points on the interior of the simplex, but they cannot be evolutionarily stable. See Figure 1.2.1 for example phase portraits. As a consequence of these results, the replicator equation can only say that languages are static, and robust when perturbed; these properties are true of human language only for short time spans. So, the replicator equation as it stands is of limited value in understanding language change.

1.3. Replicator dynamics with learning as a model for language change

In this section, we will extend the replicator equation (1.2.3) to include an abstract learning process. This extended dynamical system will be able to model spontaneous change, regular oscillations, and sensitivity as observed in actual languages.

The setup is as in Section 1.2.2, with the addition of a row-stochastic matrix Q that represents the learning process: $Q_{i,j}$ is the probability that a child of a speaker of G_i ends up speaking G_j. The diagonal entry $Q_{i,i}$ represents the learning reliability of G_i, and the off-diagonal entries represent the probabilities of making each possible learning mistake. A reasonable assumption is that Q should be diagonally dominant, meaning that it is most likely that children learn their parents' grammar correctly.

The new model, known as the *language dynamical equation* [36], is as follows:

$$
\begin{aligned}
F &= Bx. \\
\phi &= x^T Bx. \\
\dot{x}_j &= \sum_{i=1}^{n} x_i F_i Q_{i,j} - x_j \phi. \\
&= x_j(F_j Q_{j,j} - \phi) + \sum_{i \neq j} F_i x_i Q_{i,j}.
\end{aligned}
\tag{1.3.1}
$$

The second form of $\dot{x}_j$ shows the relationship to the replicator equation. The left-hand term is the same as (1.2.3) except for the learning reliability term $Q_{j,j}$. The summation represents contribution due to learning error. The original replicator equation may be recovered by imposing perfect learning, that is, by setting Q to the identity matrix. Like the replicator equation, this dynamical system has a cubic non-linearity.

The same model may be used with the interpretation that Q is a mutation process among genotypes or variants of a complex molecule or quasispecies [15, 16, 17, 31, 66, 67, 69]. In that case, the Q terms are generally treated as a small perturbation from pure replicator dynamics. In the linguistic interpretation on the other hand, the learning process itself is an object of primary interest, and learning errors may not be rare enough for Q to be interpreted as a perturbation. Furthermore, there is a correspondence between the language

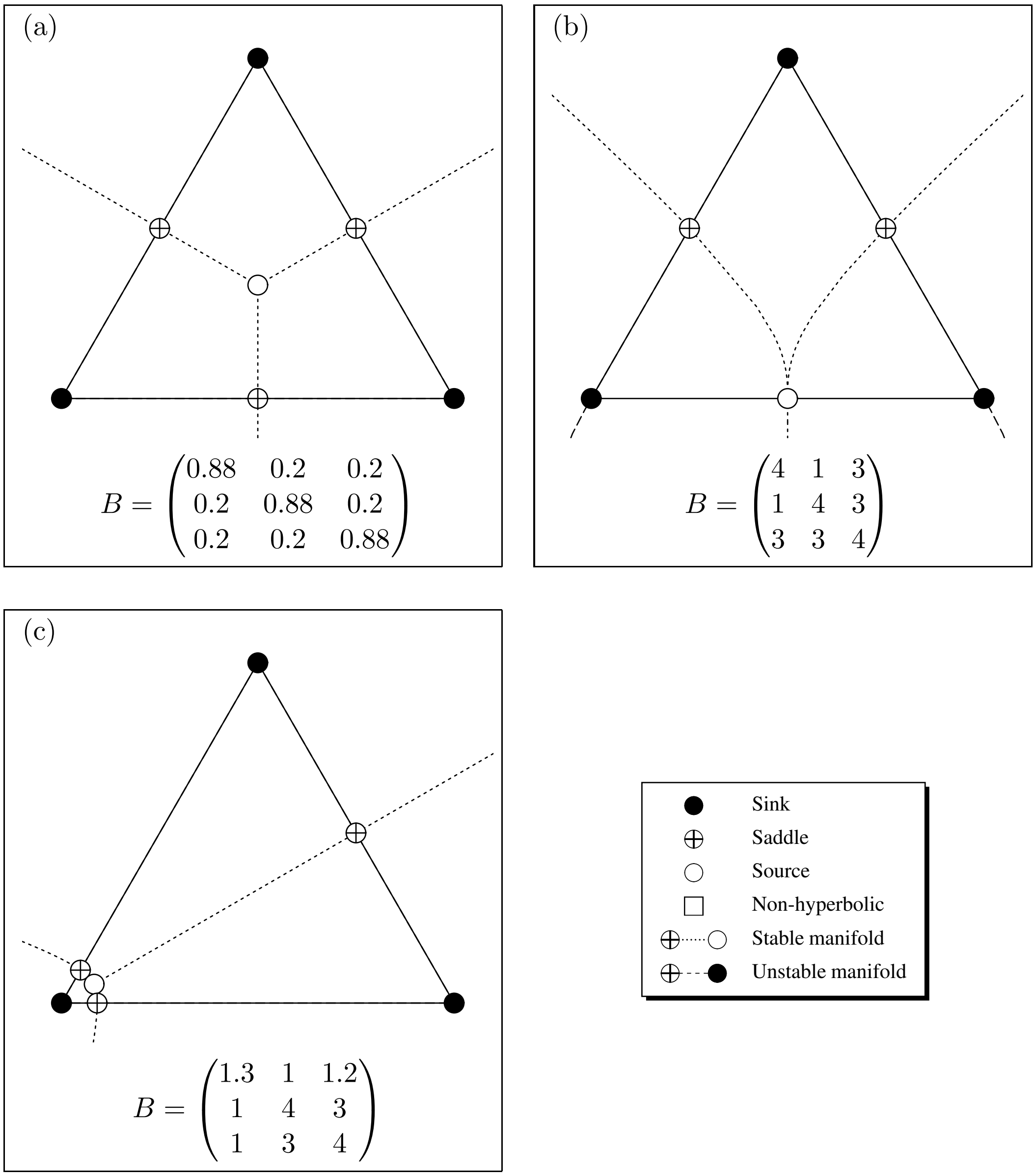


Figure 1.2.1. Example phase portraits for replicator dynamics with diagonally dominant payoff matrices. In these pictures, there are three grammars, so the phase plane is S_3, which may be represented as a triangle. The payoff matrix for each picture is as indicated. Observe that the corners are always stable sinks. The bottom left corner of each triangle is e_1, representing a population state where everyone speaks G_1. The bottom right corner is e_2, and the top is e_3.

dynamical equation and the Price equation, which is considered to be a general description of evolutionary dynamics [21, 60, 64].

The language dynamical equation shares many properties of the replicator equation. Note that $\sum_j x_j$ is always 1 as required. As before, we may define

$$M_1 = \sum_{j=1}^{n} x_j, \tag{1.3.2}$$

and compute

$$\begin{aligned}
\dot{M}_1 &= \sum_{j=1}^{n} \dot{x}_j \\
&= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} F_i x_i Q_{i,j} - \phi x_j \right) \\
&= \sum_{i=1}^{n} \left(F_i x_i \sum_{j=1}^{n} Q_{i,j} \right) - \phi \sum_j x_j \\
&= (1 - M_1)\phi,
\end{aligned}$$

which has a stable fixed value $M_1 = 1$. 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. As with the replicator equation, the phase space of the language dynamical equation is limited to the simplex S_n.

The payoff matrix B has a certain dimensionless quality, arising from the following scaling invariant. Let $\bar{B} = cB$, where c is a scalar constant, and let $\bar{x}$, $\bar{F}$, and $\bar{\phi}$ be the variables in a language dynamical equation based on $\bar{B}$. Then,

$$\begin{aligned}
\dot{\bar{x}}_j &= \sum_{i=1}^{n} \bar{x}_i \bar{F}_i Q_{i,j} - \bar{x}_j \bar{\phi} \\
&= \sum_{i=1}^{n} \bar{x}_i (e_i^T \bar{B} \bar{x}) Q_{i,j} - \bar{x}_j (\bar{x}^T \bar{B} \bar{x}) \\
&= c \left(\sum_{i=1}^{n} \bar{x}_i (e_i^T B \bar{x}) Q_{i,j} - \bar{x}_j (\bar{x}^T B \bar{x}) \right).
\end{aligned}$$

Now rescale time by setting $t = c\tau$, so that

$$\begin{aligned}
\frac{d\bar{x}_j}{d\tau} &= \frac{\dot{\bar{x}}_j}{c} \\
&= \sum_{i=1}^{n} \bar{x}_i (e_i^T B \bar{x}) Q_{i,j} - \bar{x}_j (\bar{x}^T B \bar{x}),
\end{aligned}$$

which is exactly the vector field for the language dynamical equation with payoff matrix B. Thus, if B is scaled, then a rescaling of time restores the original dynamical system.

Additionally, an "index explosion" technique allows several generalizations of (1.3.1) to be re-interpreted as instances of the same equation in higher dimensions. For example, consider bilingualism. Let (h, k) be an index representing a person who speaks G_h and G_k. A monolingual individual is associated to an index such as (h, h). We may denote by $x[(h, k)]$ the fraction of the population that speaks G_h and G_k, and by $Q[(h_1, k_1), (h_2, k_2)]$ the probability that a child growing up in a household that uses G_{h_1} and G_{k_1} ends up speaking G_{h_2} and G_{k_2}. We may now number the indices (h, k) from 1 to n^2, and use the language dynamical equation with n^2 variables to represent the dynamics. As another example, consider spatial effects, such as clustering, which may 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. Suppose that there are m counties, and we are interested in the effects of spatial locality on language dynamics. Now, an index (i, c) represents someone in county c who speaks G_i. In this case, $x[(i, c)]$ represents the part of the population in county c that speaks G_i, and $Q[(i_1, c_1), (i_2, c_2)]$ represents the probability that a parent in c_1 who speaks G_{i_1} produces a child who speaks G_{i_2} and moves to county c_2. Again, the result is a new instance of the language dynamical equation with nm variables. Of course, such index explosions could also be combined to model multilingualism with spatial effects, for example. Since index explosion can turn so many extensions into higher dimensional instances of the original model, we will focus on the original model for now and leave those extensions to future work.

At this point, it should be clear that this model can get out of hand in a hurry. Since the number of parameters is roughly the square of the number of variables, even a relatively small number of grammars results in a model that is practically impossible to analyze in full. Thus to make progress, it will be helpful to try one of the following techniques.

Symmetry: We can keep the number of variables arbitrary, but impose some sort of symmetry, so that the number of parameters is manageable.

Low dimensions: We can keep the parameters fully general, but limit the number of dimensions to 2 or 3 so that the resulting dynamical system is manageable.

Parameter independence: Some properties of the dynamical system may be independent of the parameters under certain circumstances.

Several mixtures of these techniques are used in the rest of this dissertation to analyze the language dynamical equation.

The model can also be extended to include genetic variation, which opens the door to analysis of many more situations. For studying language change on geological time scales, it is necessary to include evolutionary change to the language faculty itself, and therefore variation among UGs. As noted earlier, it is not clear how much variation is present in human UG, and this model can be used to formulate questions about which UGs can coexist, and which exclude one another. By addressing these issues, the multi-UG model can be used to understand the evolution of language.

The multi-UG extension is straightforward: We assume that there are N possible UGs, denoted $U_1, U_2, \ldots, U_N$, that collectively admit n possible grammars $G_1, G_2, \ldots, G_n$. The payoff matrix B is the same as before, but the learning matrix Q must be augmented by a third index: $Q_{i,j,K}$ is now the probability that a child of a speaker of G_i ends up speaking G_j, given that both have U_K. The population will be represented by variables $x_{i,K}$ that represent the fraction of the population with U_K that speaks G_i. We will also be interested in the fraction of the population with U_K, denoted y_K,

$$y_K = \sum_{j=1}^{n} x_{j,K}, \tag{1.3.3}$$

and the fraction speaking G_j,

$$w_j = \sum_{K=1}^{N} x_{j,K}. \tag{1.3.4}$$

The language dynamical equation for multiple universal grammars is

$$\begin{aligned} F &= Bw. \\ \phi &= w^T B w. \\ \dot{x}_{j,K} &= \sum_{i=1}^{n} F_i x_{i,K} Q_{i,j,K} - \phi x_{j,K}. \end{aligned} \tag{1.3.5}$$

A few remarks are in order. First, index explosion could be used to turn a multi-UG language dynamical equation into a high-dimensional case of the single-UG equation. However, enough interesting results may be obtained from the multi-UG equation that it is worth the trouble to analyze separately and in detail. Second, it is possible that some U_K does not admit a particular G_j. To handle this case, the corresponding population variable $x_{j,K}$ is frozen at 0, and Q must be consistent with this restriction: For all i, it must be true that $Q_{i,j,K} = 0$ because it is impossible for someone with U_K to acquire G_j, even by mistake.

Finally, the multi-UG equation represents a theoretical extension to population models under game dynamics. Consider a communication game among universal grammars. The difficulty is in defining a payoff function $P(U_H, U_K)$ that gives the payoff of U_H playing against U_K. Such a payoff function cannot be defined, because the reward in a communication game comes from the players' strategies, their grammars in this case, and not directly from a UG, which is a strategy for selecting a strategy. Hence, a UG might be called a *metastrategy,* and a competition among UGs might be called a *metastrategy game.* Since UGs do not directly determine payoff, the notions of Nash equilibrium and ESS cannot be directly applied to metastrategy games, and other methods of analysis must be found.

1.4. Outline

The rest of the dissertation is organized as follows. Chapter 2 extends some of the calculations in [36] concerning the language dynamical equation in an arbitrary number of dimensions, but with maximum symmetry. The results include a complete characterization of the possible behavior of the system under full symmetry. All 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. Although this form of the system has fairly static behavior, it can still be used to understand language change: If the environment changes, due to contact with other languages for example, then the learning process may also change, thereby destabilizing an equilibrium population. (The replicator equation lacks this interpretation as the only parameters represent payoff from communication ability, which seems less subject to change.)

Chapter 3 describes some more exciting behavior for the language dynamical equation: oscillations, period doubling, and chaos. Remember that part of the goal of this research is to qualitatively reproduce different types of observed language change through a dynamical system. This chapter shows that the model can indeed mimic spontaneous fluctuations, regular changes, unpredictability, and sensitivity.

The second part of the dissertation is concerned with the extended language dynamical equation that allows for multiple universal grammars. It can be used to study genetic variation in UG and potentially to answer questions about the genetic history and evolution of the language faculty. Chapter 4 is about competition between UGs in some low-dimensional cases, specifically, when one of the UGs in question admits only one grammar. The chapter includes examples for competitive exclusion and stable coexistence of different UGs. We will analyze conditions for single-grammar UGs to out-compete multi-grammar UGs and vice versa. An interesting finding is that multi-grammar UGs can resist invasion by single-grammar UGs if learning is more accurate. In other words, accurate learning stabilizes UGs that admit large numbers of candidate grammars.

In Chapter 5, we will continue the discussion of competition between UGs, focusing this time on a three dimensional case: Each UG admits two grammars. This case is distinctly non-trivial, and several exact results are presented for parameters with low symmetry, as well as a general result concerning when the two UGs are each stable against invasion by the other. Chapters 4 and 5 form the beginning stages of analysis of metastrategy games.

Finally, in Chapter 6, we will draw some conclusions and sketch some possibilities for future work.

Chapter 2

Bifurcations of the Fully Symmetric Language Dynamical Equation

Contents

2.1. Introduction

As described in Chapter 1, any number of mathematical frameworks have been proposed for modeling the evolution of languages. This dissertation is concerned with the model

[†]The bulk of this chapter has been published as [48], Copyright © Springer-Verlag 2002.

described in [36], 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 chapter extends the analysis in [36] and adds several new results. The focus of this chapter is to provide a complete bifurcation analysis of the language dynamical equation in a special case where the number of grammars is arbitrary, but the parameters are chosen to make the dynamical system highly symmetric, and therefore tractable.

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 main result of this chapter is a description of how learning reliability affects the stability of these two kinds of equilibria. We will rigorously find all fixed points, determine their stabilities, and prove that all populations tend to some fixed point. We will 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. The bifurcation analysis presented here provides a mathematical description of how the transition from incoherence to coherence takes place.

In its fully general form, the language dynamical equation is a system of non-linear ordinary differential equations (ODEs) 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 2.2 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 [36]. Section 2.3 gives an outline of the bifurcation scenario and pictures from the three-grammar case. In Section 2.4, we determine the locations of all fixed points and the parameter values for which they exist. Section 2.5 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 2.6, including the $\mathcal{S}_n$ transcritical bifurcation in which the incoherent equilibrium reverses stability.

Further analysis in Section 2.7 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.

2.2. Parameter settings

The fully general model (1.3.1) is too complex to analyze without some simplifying assumptions. Following Komarova *et al.* [36], we will constrain the B 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.

Consider an idealized abstract communication game. 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 payoff matrix by

$$B_{i,j} = \left(\alpha A_{i,j} + (1-\alpha)A_{j,i}\right). \tag{2.2.1}$$

That is, payoff 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. 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

$$B_{i,j} = \frac{A_{i,j} + A_{j,i}}{2}, \text{ or } B = \frac{1}{2}(A + A^T). \tag{2.2.2}$$

We will also assume that all grammars in question are unambiguous, so $A_{i,i} = 1$. For the rest of this chapter, we will assume the following form for A and Q:

$$A = \begin{pmatrix} 1 & a & \cdots & a \\ a & 1 & \cdots & a \\ \vdots & \vdots & \ddots & \vdots \\ a & a & \cdots & 1 \end{pmatrix}, \tag{2.2.3}$$

$$Q = \begin{pmatrix} q & u & \cdots & u \\ u & q & \cdots & u \\ \vdots & \vdots & \ddots & \vdots \\ u & u & \cdots & q \end{pmatrix}, \quad \text{where } u = \frac{1-q}{n-1}. \tag{2.2.4}$$

Since $A = A^T$, it follows that $B = A$. 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:

$$M_k = \sum_{j=1}^{n} x_j^k. \tag{2.2.5}$$

Simplifying the original form of the language dynamical equation (1.3.1) and incorporating the restrictions on A and Q yields the following two expressions for the dynamics:

$$\dot{x}_j = (1-a)\left(-x_j^3 + qx_j^2 + (u - x_j)\sum_{i \neq j} x_i^2\right) - au\,(nx_j - 1) \tag{2.2.6a}$$

$$= (1-a)\big((q-u)x_j^2 + uM_2 - x_jM_2\big) - aunx_j + au. \tag{2.2.6b}$$

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 (2.2.6) as the *fully symmetric language dynamical equation*, and the rest of this chapter is concerned with this restricted form of (1.3.1).

2.3. 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 2.3.1. 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 2.3.2. 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 2.3.3. 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.

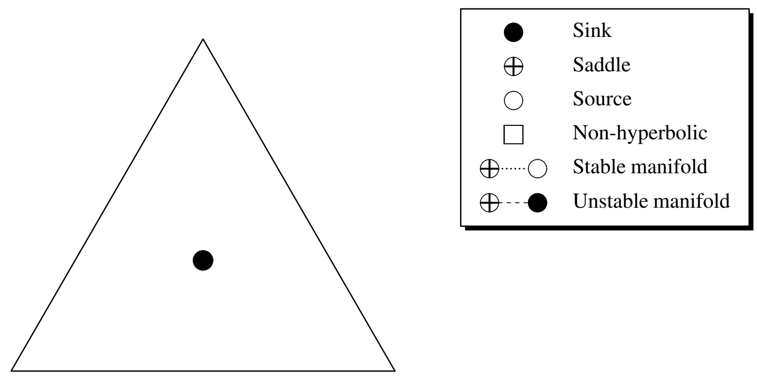


Figure 2.3.1. Phase portrait with $a = 0.5$, $q = 0.85$.

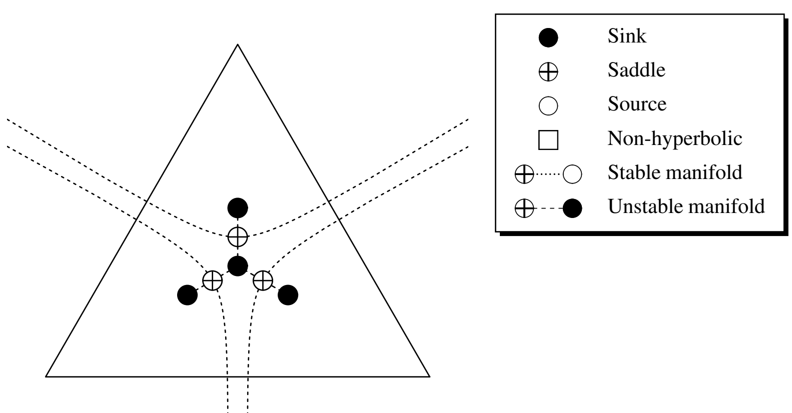


Figure 2.3.2. Phase portrait with $a = 0.5$, $q = 0.8575$.

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 provides enough illustration to give the reader some intuition for the analysis that follows.

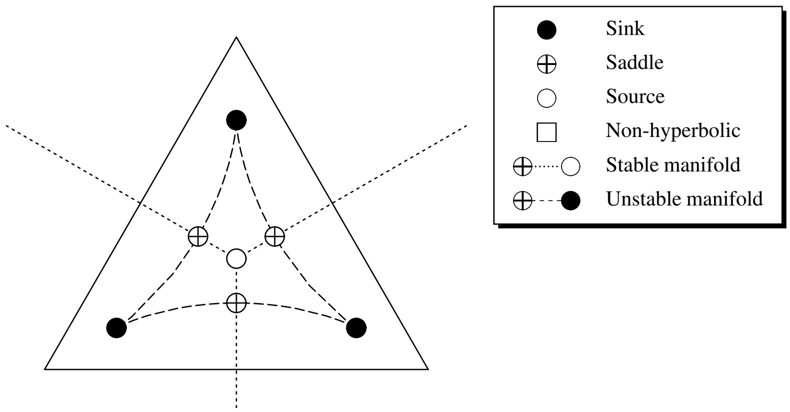


Figure 2.3.3. Phase portrait with $a = 0.5$, $q = 0.9$.

2.4. 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 (2.2.6) will have some symmetric form. In particular, we make the assumption that at fixed points, m grammars will share the majority of the population equally, and the rest will split the remainder equally.

Proposition 2.4.1. *Every fixed point $\bar{x}$ of (2.2.6) 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

$$(1 - a)\left((q - u)Z^2 + uM_2 - ZM_2\right) - aunZ + au = 0. \tag{2.4.1}$$

This polynomial, which comes from (2.2.6b), 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. □

We define $X^{(m)}$ and $Y^{(m)}$ to be the roots of (2.4.1), 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.*[1] 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.

[1] Komarova *et al.* [36] refers to these as *m-grammar fixed points* because in most cases, the fixed points represent states of the population in which most of the population speak one of m of the grammars, and only a small fraction speak the others. This convention is confusing in the case $m = 1$ because for a range of q, both the 1-up and $(n-1)$-up fixed points correspond to states where a single grammar dominates. So, the m-up convention is used here in the hope that it will be less confusing.

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 (2.2.6a). 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 2.4.1 into (2.2.6b). 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 the following cubic equation:

$$\begin{aligned} &\left(-\frac{(1-a)mn}{n-m}\right)Z^3 \\ &+\left(\frac{(1-a)\left(m+n-3mn+2mnq-n^2q\right)}{(n-m)(n-1)}\right)Z^2 \\ &-\left(\frac{n-1-2m(q-1)+a\left(1-n+n^2+m(n+2)(q-1)-n^2q\right)}{(n-m)(n-1)}\right)Z \\ &-\left(\frac{(a(1+m-n)-1)(1-q)}{(n-1)(n-m)}\right)=0. \end{aligned} \tag{2.4.2}$$

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

$$\begin{aligned} &(a-1)m(n-1)Z^2 \\ &+(a-1)(1+2m(q-1)-nq)Z \\ &-(a(1-m-n)-1)(q-1)=0. \end{aligned} \tag{2.4.3}$$

The roots are found with the quadratic formula, yielding

$$Z_{\pm}^{(m)} = -\frac{1-2m+2mq-nq}{2m(n-1)} \pm \frac{\sqrt{d}}{2m(1-a)(n-1)}, \tag{2.4.4}$$

where the discriminant d is given by

$$\begin{aligned} d =&(1-a)\Big(4m(n-1)(1-q)(a+am-an-1) \\ &+(1-a)(1-2m(1-q)-nq)^2\Big). \end{aligned} \tag{2.4.5}$$

The cubic equation (2.4.2) was set up to look for values of Z such that some fixed point has m elements equal to Z. Therefore,

$$\begin{aligned} X^{(m)} &= Z_+^{(m)}, \\ Y^{(m)} &= Z_-^{(n-m)} = \frac{1 - mZ_+^{(m)}}{n-m}. \end{aligned} \tag{2.4.6}$$

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

$$\begin{aligned} \hat{q}_m = \frac{1}{(a-1)(n-2m)^2}\Big(&2m(n-m)(2+a(n-3)) + (a-1)n \\ &- 2(n-1)\sqrt{(1+a(m-1))(1+a(n-m-1))m(n-m)}\Big). \end{aligned} \tag{2.4.7}$$

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 2.4.1 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.

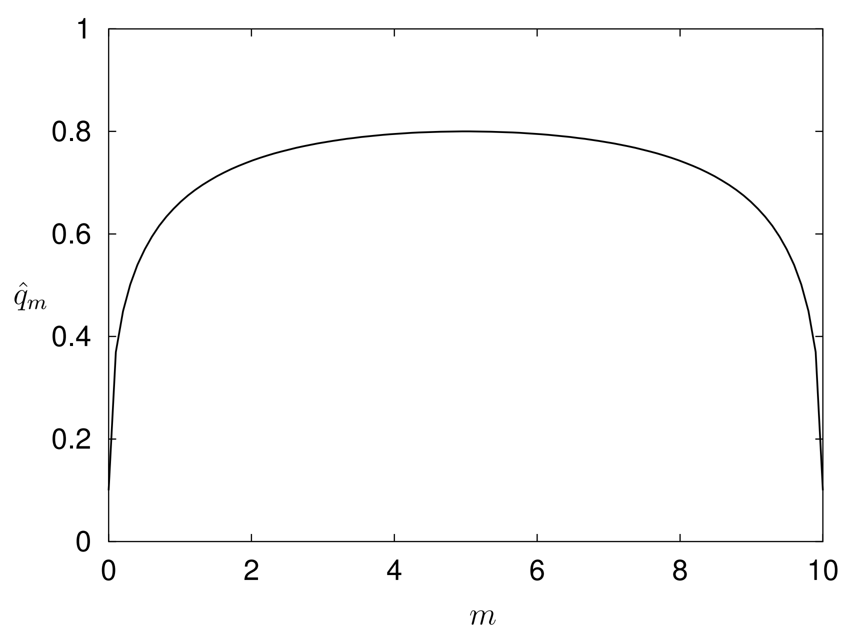


Figure 2.4.1. Plot of $\hat{q}_m$, with $a = 0.2$ and $n = 10$. The m-up fixed points do not exist until $q > \hat{q}_m$.

More rigorously, we can make the substitution $m = n/2 + h$. After some simplification, (2.4.7) becomes

$$\hat{q}_m = \frac{2 + a(n-3)}{2(1-a)} + \frac{(n-1)n\left(1 - a + \frac{an}{2}\right)}{4(1-a)} g(h), \tag{2.4.8}$$

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 when $a = 1$, in which case all the languages are identical, or $a = 1/(1-n) < 0$; 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 2.4.2. *The function $g(h)$ has a global maximum at $h = 0$ and is concave down for $-n/2 \leq h \leq /2$.*

Proof. Observe that g, being a quotient of two analytic functions, is meromorphic near $h = 0$, so $g(0) = \lim_{h\to 0} g(h)$ does exist in the sense of complex numbers, although it may be infinity. To prove that $g(0)$ is in fact finite, note that for small h, we can expand $\sqrt{1 - \beta h^2}$ into the Taylor series $1 + \beta h^2/2 + \beta h^4/8 + O\left(()\, h^6\right)$. Thus, the numerator of g is $-(\beta_1 + \beta_2)h^2/2 - (\beta_1 - \beta_2)^2 h^4/8 + O\left(()\, h^6\right)$ which means that $g(h) = -(\beta_1+\beta_2)/2 - (\beta_1 - \beta_2)^2 h^2/8 + O\left(()\, h^4\right)$ which is bounded for small h. 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:

$$g'(h) = \frac{(\beta_1 + \beta_2)h^2 - 2 + 2\sqrt{(1-\beta_1 h^2)(1-\beta_2 h^2)}}{\sqrt{(1-\beta_1 h^2)(1-\beta_2 h^2)}h^3}.$$

If g' is to be zero, its numerator must be zero, which implies

$$4\left(1 - (\beta_1+\beta_2)h^2 + \beta_1\beta_2 h^4\right) = 4 - 4(\beta_1 + \beta_2)h^2 + (\beta_1+\beta_2)^2 h^4.$$

After canceling terms, the equation reduces to

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

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 2.4.1. In particular, $\hat{q}_m$ has a global maximum at $m = n/2$, given by

$$\hat{q}_{\max} = \hat{q}_m\big|_{m=\frac{n}{2}} = \frac{1+n+a(n^2-n-1)}{n(2-2a+an)}, \tag{2.4.9}$$

after much simplification.

2.5. 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 (2.2.6) at the various fixed points, derive expressions for its eigenvalues, and determine their multiplicities. In Section 2.6, 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.[2]

The Jacobian matrix for (2.2.6) has entries of two types:

$$\frac{\partial \dot{x}_i}{\partial x_i} = (1-a)\left(2x_i\left(q - x_i\right) - M_2\right) - aun, \tag{2.5.1}$$

and for $j \neq i$:

$$\frac{\partial \dot{x}_i}{\partial x_j} = 2(1-a)\left(u - x_i\right)x_j. \tag{2.5.2}$$

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

[2] It is tempting to start with the simplex constraint $\sum_j x_j = 1$ and take derivatives with respect to t and x_i to arrive at the identity $\sum_j \partial \dot{x}_j / \partial x_i = 0$; however, the constraint removes one degree of freedom, so the variables x_i are no longer independent. For this reason, this identity cannot be used in the following analysis. Other identities do hold, however, that simplify the calculation of the eigenvalues of the Jacobian.

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 (2.2.6) at an m-up fixed point with the first m entries equal to $X^{(m)}$ takes the form

$$
J = \left(\begin{array}{cccc|cccc}
c_1 & c_3 & \cdots & c_3 & c_4 & c_4 & \cdots & c_4 \\
c_3 & c_1 & \cdots & c_3 & c_4 & c_4 & \cdots & c_4 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
c_3 & c_3 & \cdots & c_1 & c_4 & c_4 & \cdots & c_4 \\
\hline
c_6 & c_6 & \cdots & c_6 & c_2 & c_5 & \cdots & c_5 \\
c_6 & c_6 & \cdots & c_6 & c_5 & c_2 & \cdots & c_5 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
c_6 & c_6 & \cdots & c_6 & c_5 & c_5 & \cdots & c_2
\end{array}\right). \tag{2.5.3}
$$

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:

$$
\begin{aligned}
c_1 r + (m-1)c_3 r + (n-m)c_4 s &= \lambda r, \\
m c_6 r + c_2 s + (n-m-1)c_5 s &= \lambda s.
\end{aligned} \tag{2.5.4}
$$

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

$$mr + (n-m)s = 0. \tag{2.5.5}$$

Using (2.5.5) to solve for s in terms of r and substituting that expression for s in the first equation of (2.5.4) yields

$$\lambda_1 = c_1 + (m-1)c_3 - mc_4. \tag{2.5.6}$$

Alternatively, we could solve for r in terms of s and use the second equation in (2.5.4), which gives the solution in a different form:

$$\lambda_1 = c_2 + (n-m-1)c_5 - (n-m)c_6. \tag{2.5.7}$$

A computation confirms that both expressions for λ_1 are equal. A particular eigenvector v_1 corresponding to λ_1 may be found from (2.5.5), for example, by setting $r = (n-m)$ and $s = -m$.

A second eigenvalue may be determined by computing the trace of the system (2.5.4) and subtracting λ_1. The result is

$$\begin{aligned} \lambda_0 &= c_2 + (n-m-1)c_5 + mc_4 \\ &= c_1 + (m-1)c_3 + (n-m)c_6, \end{aligned} \tag{2.5.8}$$

and again, a calculation confirms that both expressions are the same. 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[3]

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

are eigenvectors such that

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

Likewise, the $n-m-1$ vectors

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

are eigenvectors with

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

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 $n-1$.

[3] The notation e_k means the k-th standard basis vector of $\mathbf{R}^n$.

2.6. Bifurcations of fixed points

In Section 2.5, 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.

2.6.1. 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 2.6.1. *For an m-up fixed point where $m > n/2$, 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 $0 < m \leq n/2$, then $\lambda_2 \geq 0$ for $q = \hat{q}_m$ and is strictly positive for $q > \hat{q}_m$.*

Proof. We look for the special value of q such that $\lambda_2 = 0$ by solving a pair of quadratic equations: The first is (2.4.3) with Z replaced by X, which constrains X to be either $X^{(m)}$ or $Y^{(n-m)}$. The second line, (2.6.1b), 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:

$$\begin{aligned}(a-1)m(n-1)X^2 \\ +(a-1)(1+2m(q-1)-nq)X \\ -(a(1-m-n)-1)(q-1)=0,\end{aligned} \tag{2.6.1a}$$

$$\begin{aligned}&\left(-\frac{(1-a)mn}{n-m}\right)X^2 \\ &+\left(\frac{2(1-a)n(-1+m-mq+nq)}{(n-m)(n-1)}\right)X \\ &+\left(-\frac{1-a}{n-m}-\frac{an(1-q)}{n-1}\right)=0.\end{aligned} \tag{2.6.1b}$$

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 (2.6.1a) by $-n$ and (2.6.1b) by $(n-1)(n-m)$ and adding the two results together yields, after much simplification:

$$-(1-a)(nq-1)(nX-1) = 0. \tag{2.6.2}$$

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

$$q = \frac{1+n+a\left(n^2-n-1\right)}{n(2+a(n-2))} = \hat{q}_{\max}. \tag{2.6.3}$$

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 (2.4.4) and (2.4.6) that

$$\begin{aligned} X^{(m)} &= -\frac{1-2m+2mq-nq}{2m(n-1)} + \frac{\sqrt{d}}{2m(1-a)(n-1)} \\ &\geq \frac{1-q}{n-1} + \frac{nq-1}{2m(n-1)}. \end{aligned}$$

The inequality is obtained by striking the second term and expanding the first. 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

$$\begin{aligned} X^{(m)} &> \frac{1-q}{n-1} + \frac{nq-1}{2m(n-1)}\bigg|_{m=\frac{n}{2}} \\ &= \frac{1}{n}. \end{aligned}$$

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 (2.4.3) 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 2.6.2. *For an m-up fixed point where $m > n/2$, 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 $\lambda_3 \leq 0$ for $q = \hat{q}_m$ and is strictly negative for $q > \hat{q}_m$.*

Proof. 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 2.6.3. *For an m-up fixed point where $m > n/2$, 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

$$\begin{aligned} &(a-1)m(n-1)X^2 \\ &+ (a-1)(1 + 2m(q-1) - nq)X \\ &- (a(1-m-n) - 1)(q-1) = 0, \end{aligned} \tag{2.6.4a}$$

$$\left(\frac{-3(1-a)mn}{n-m}\right)X^2 + \left(-\frac{2(1-a)(m+n-3mn+2mnq-n^2q)}{(n-m)(n-1)}\right)X + \left(\frac{1-n-2m(1-q)-a(1-n+n^2-m(n+2)(1-q)-n^2q)}{(n-m)(n-1)}\right) = 0. \tag{2.6.4b}$$

The first one, (2.6.4a), is the same as (2.6.1a) and constrains X to be either $X^{(m)}$ or $Y^{(n-m)}$. The second one, (2.6.4b), is an expanded form of $\lambda_1 = 0$. The linear combination of $-3n$ times (2.6.4a) plus $(n-1)(m-n)$ times (2.6.4b) 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. □

2.6.2. 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

$$q = \frac{1+n+a\left(n^2-n-1\right)}{n(2+a(n-2))} = \hat{q}_{\max}. \tag{2.6.5}$$

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.

2.6.3. 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 $q = \hat{q}_{\max}$. These are saddle-node bifurcations associated with the sign change in λ_1. At $q = \hat{q}_{\max}$, the m-up fixed points for $m > n/2$ all collide with the uniform fixed point in the center of the simplex. This mass collision of fixed points is associated with sign changes in λ_1, λ_2, and λ_3. 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 [6].)

2.7. Other properties of the vector field

The vector field given by (2.2.6) can be written as the gradient of a function $V(x)$ plus an additional term. A number of well-known proofs [26, 30] 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 2.7.1. *Define the function $V(x)$ as follows:*

$$V = \frac{1}{3}(1-a)(q-u)M_3 - \frac{1}{4}(1-a)M_2^2 - \frac{1}{2}aunM_2 + auM_1. \tag{2.7.1}$$

If $x(t)$ is a trajectory of (2.2.6) 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 (2.2.6) as possible:

$$\frac{\partial V}{\partial x_j} = (1-a)(q-u)x_j^2 - (1-a)M_2 x_j - aunx_j + au.$$

Thus, (2.2.6) may be re-written as

$$\dot{x}_j = \frac{\partial V}{\partial x_j} + (1-a)uM_2, \tag{2.7.2a}$$

or in vector notation:

$$\dot{x} = DV + (1-a)uM_2\mathbf{1}. \tag{2.7.2b}$$

Computing the time derivative of V and using (2.7.2b) to substitute for DV yields

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

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

$$\dot{V} = \|\dot{x}\|^2. \tag{2.7.3}$$

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 2.7.2. *The ODE given by (2.2.6) 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 2.7.3. *All orbits of* (2.2.6) *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 (2.7.3), $\dot{x} \to 0$, which is only possible if the orbit converges to a fixed point. □

2.8. 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 noisy learning environments. Consider for example the transition from Old English to Middle English. One theory, described in [45], is that part of the change was due to the influence of Scandinavian invaders from the eighth to eleventh centuries. Their language, Old Norse, was similar to Old English, both being Germanic languages. For example, both languages used similar case endings on nouns, as illustrated in their words for *stone* [45, p. 11]:

	Old English		**Old Norse**	
	Singular	Plural	Singular	Plural
Nominative	stān	stānas	steinn	steinar
Accusative	stān	stānas	stein	steina
Genitive	stānes	stāna	steins	steina
Dative	stāne	stānum	steini	steinum

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.) 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 directly 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, [6] for another instance of this bifurcation and [20] for a relevant theorem.) An alternative would be to explore parameter settings which have a smaller symmetry group, for example, cyclic or dihedral symmetry.

Chapter 3

Chaos and Oscillations in Language Dynamics

Contents

3.1. Introduction

Over historical time scales, languages change dramatically and unpredictably by accumulation of small changes and by interaction with other languages. The results of Chapter 2 can be interpreted in a manner consistent with language change due to contact or other external influences. However, when left to itself, the language dynamical equation with full symmetry will converge to a fixed point, thereby failing to capture the phenomenon of spontaneous language change. This chapter shows that simple instances of the language dynamical equation can display complex limit cycles and chaos, thereby mimicking complicated and unpredictable changes of languages over time.

The payoff matrix B is assumed to represent some sort of communication game. Its entries may include effects such as the benefit of correct communication, cost of ambiguity, and so forth. A natural assumption is that people communicate best with others who have the same grammar. Consequently, B is diagonally dominant, which implies that each grammar is a strict Nash equilibrium: If the whole population speaks the same language, then no individual can receive a higher payoff by switching to another language. With perfect learning, each language would then be an evolutionarily stable equilibrium [47], but imperfect learning can lead to chaotic oscillations.

Section 3.2 describes particular parameter settings that produce regular oscillations and a spiral sink in two dimensions (that is, three grammars). The spiral sink can be converted to a higher-dimensional limit cycle, and with the right choice of parameters, this new limit cycle can undergo period doubling bifurcations that lead to chaos. The mechanism underlying this instance of chaos seems to be a variation of Smale's horseshoe [14, 26]. A Poincaré map illustrating possible horseshoes is given in Section 3.3. The chaotic orbits in this example are based on a sort of escape-and-return mechanism that is potentially useful for modeling other phenomena of language change. In particular, any number of "modules" can be joined by this escape-and-return mechanism, resulting in arbitrarily complex limit cycles. The connection to the modular structure of language is discussed in Section 3.4.

In Section 3.5, we conclude by drawing parallels between the example regular and chaotic oscillations and the morphological type cycle, spontaneous fluctuations, and sensitivity observed in actual languages.

3.2. Limit cycles and chaos

As a specific example, let us consider the following payoff matrix

$$B = \begin{pmatrix} 0.88 & 0.2 & 0.2 \\ 0.2 & 0.88 & 0.2 \\ 0.2 & 0.2 & 0.88 \end{pmatrix}. \tag{3.2.1}$$

There are only three grammars, G_1, G_2 and G_3. The highest payoff is given for communication among identical grammars. All grammars are equally good, and all off-diagonal values are identical.[1]

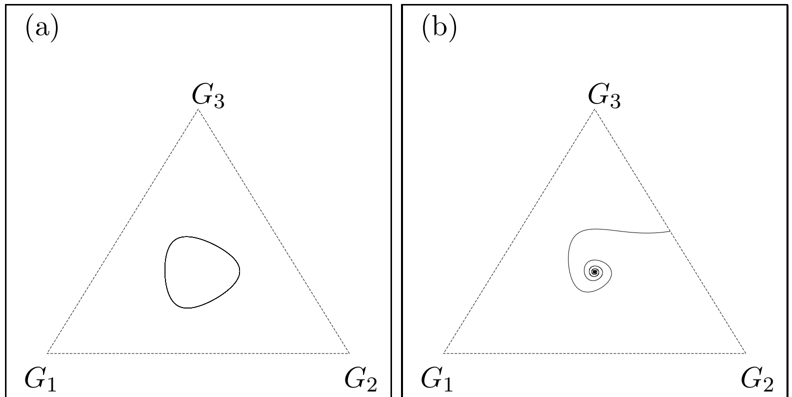
For perfect learning, this payoff matrix leads to very simple dynamics: There are stable fixed points in the corners of the simplex and no stable fixed points in the interior, as in Figure 1.2.1a. All trajectories converge to one of the three stable equilibria, so at equilibrium, the whole population speaks the same language.

Imperfect learning, however, can induce very different behavior. For example, let us consider the following learning matrix in conjunction with (3.2.1):

$$Q = \begin{pmatrix} 0.79 & 0.2 & 0.01 \\ 0.01 & 0.79 & 0.2 \\ 0.2 & 0.01 & 0.79 \end{pmatrix}.$$

For each grammar, the most likely outcome of the learning process is the correct grammar. For each grammar, there is one second most likely outcome, and the errors are structured so as to produce rotation in the dynamical system. As shown in Figure 3.2.1a, these parameters produce stable oscillations, as learning errors in the subpopulation speaking G_1 feed into G_2, and G_2 feeds into G_3, and G_3 feeds back into G_1. The population never settles for one language, but continuously oscillates among all three languages. We note that a stable

[1] The choice of 0.88 as opposed to 1 for the diagonal entries of B is essentially an accident. Recall that B is dimensionless, and could be rescaled without changing the dynamics. See Section 1.3.



$$Q = \begin{pmatrix} 0.79 & 0.2 & 0.01 \\ 0.01 & 0.79 & 0.2 \\ 0.2 & 0.01 & 0.79 \end{pmatrix} \qquad Q = \begin{pmatrix} 0.76 & 0.2 & 0.04 \\ 0.04 & 0.76 & 0.2 \\ 0.2 & 0.04 & 0.76 \end{pmatrix}$$

$$B = \begin{pmatrix} 0.88 & 0.2 & 0.2 \\ 0.2 & 0.88 & 0.2 \\ 0.2 & 0.2 & 0.88 \end{pmatrix}$$

Figure 3.2.1. **(a)** A stable limit cycle. The counterclockwise oscillations are caused by the learning algorithm. Errors from children learning G_1 feed into G_2, and G_2 feeds into G_3, and G_3 feeds back into G_1. **(b)** A spiral sink that results from the limit cycle in (a) when the Q matrix is changed as shown. The B matrix is common to both figures.

limit cycle, as shown in Figure 3.2.1a, is not possible for any 3-variable replicator equation [31, Theorem 4.2.1 and Section 7.5]. If learning becomes less accurate, then the limit cycle breaks down, resulting in a spiral sink, as shown in Figure 3.2.1b.

This spiral opens the way for more complex behavior, as it can be used to construct a period doubling cascade similar to Šilnikov's mechanism [26, 75, Section 6.5]. We add two more grammars, and set up the learning matrix so that the spiral sink is unstable in the new dimensions. We fix B as follows

$$B = \begin{pmatrix} 0.88 & 0.2 & 0.2 & 0 & 0.3 \\ 0.2 & 0.88 & 0.2 & 0 & 0.3 \\ 0.2 & 0.2 & 0.88 & 0 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.88 & 0 \\ 0 & 0 & 0 & 0.3 & 0.88 \end{pmatrix}. \tag{3.2.2}$$

Thus, we consider 5 languages, each of which is a strict Nash equilibrium. For perfect learning, there would again be convergence to the homogeneous states where all individuals speak the same language. Instead of perfect learning, let us consider a one-parameter family of Q matrices

$$Q = \begin{pmatrix} 0.75 & 0.2 & 0.01 & 0.04 & 0 \\ 0.01 & 0.75 & 0.2 & 0.04 & 0 \\ 0.2 & 0.01 & 0.75 & 0.04 & 0 \\ 0 & 0 & 0 & \mu & 1-\mu \\ 1-\mu & 0 & 0 & 0 & \mu \end{pmatrix}. \tag{3.2.3}$$

The parameter μ denotes the learning accuracy of grammars G_4 and G_5. For appropriate choices of μ, trajectories can escape from the middle of the G_1, G_2, G_3 spiral. The population slowly leaks into G_4, from there into G_5, and then back into G_1 to return to the spiral. Varying μ changes how accurate the learning algorithm is for G_4 and G_5, and alters how trajectories escape the spiral. The result is, initially, a more complex limit cycle in four dimensions. It can undergo period-doubling bifurcations that lead to chaos, as shown in Figure 3.2.2 [26, 46]. Note that the most interesting behavior is observed for μ values just below the learning accuracy of G_1, G_2 and G_3 which is set at 0.75.

The chaotic trajectory of Figure 3.2.2d is depicted in more detail in Figure 3.2.3, which shows the most abundant grammar at any given time. The oscillation between G_1, G_2, and G_3 is clearly visible, with irregular interruptions by the escape mechanism via G_4. The return mechanism via G_5 can be seen in Figure 3.2.3b, where each increase in x_1 is preceded by a spike in x_5.

It is known that the game dynamical equation for four or more dimensions can generate limit cycles and chaos [31, 53, 67]. Sometimes adding extremely small mutation terms (near 10^{-5}) preserves chaos, but for larger mutation rates, chaos often reverts to stable limit cycles [69]. In the present example, chaos is not caused by game dynamics, but by learning errors, which are similar to mutational processes.

3.3. The mechanism of chaos

The mechanism at work appears to be a combination of stretching and twisting, which can be seen by analyzing the three-dimensional Poincaré map obtained by taking all points with $x_4 = 0.19$. A partial return map for this Poincaré section shown in Figure 3.3.1. The initial condition is represented by the multi-colored metallic pyramid. The return map stretches it, and curls it into the spiral as shown. It appears that this map contains a variation of Smale's horseshoe [14, 26] that is responsible for the existence of chaos in this example.

3.4. Modules and clusters

Human language has a modular architecture, which contains phonetic, syntactic, and semantic subsystems that interact in complex ways [34]. Language change also tends to happen in a modular manner. For example, borrowed words can lead to the development of new phonetic rules that lead over time to new morphemes and eventually new syntax. (See Chapter 11 of [74] for a discussion of such changes in Turkish and Armenian.) The language dynamical equation can also exhibit changes with a modular character. The spiral-and-escape mechanism underlying the chaos in Figure 3.2.2 can be used to join spirals

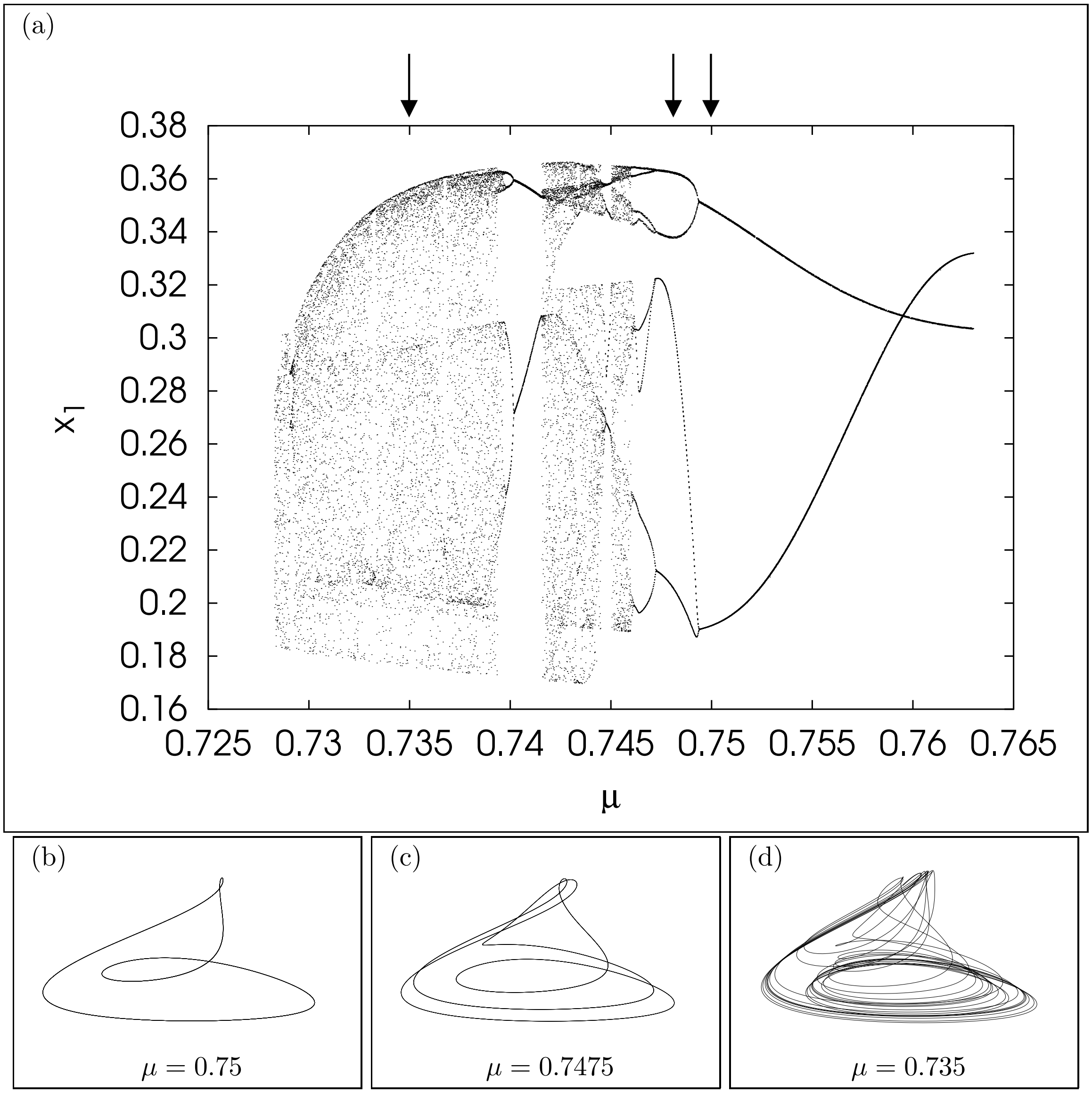


Figure 3.2.2. **(a)** Cascade diagram. The horizontal axis shows a range of values of μ. For each value of μ, an orbit is traced. Each time x_4 crosses 0.19, a dot is plotted at (μ, x_1). Sample orbits are drawn for the three values of μ indicated by arrows. **(b)** When $\mu = 0.75$, there are two dots representing two extremes of this stable limit cycle. **(c)** When $\mu = 0.7475$, there are twice as many dots because the limit cycle has undergone a period doubling bifurcation. **(d)** For $\mu = 0.735$, the orbit appears to be chaotic.

together into complex stable limit cycles, as shown in Figure 3.4.1. That figure illustrates a case of two spirals of three grammars each linked together via two more grammars, requiring a total of eight grammars. A population in the left-hand part of the picture speaks primarily G_1, G_2, and G_3. Features common to those grammars will linger until the escape mechanism kicks in. Then, the population will move via G_4 to the right-hand part of the

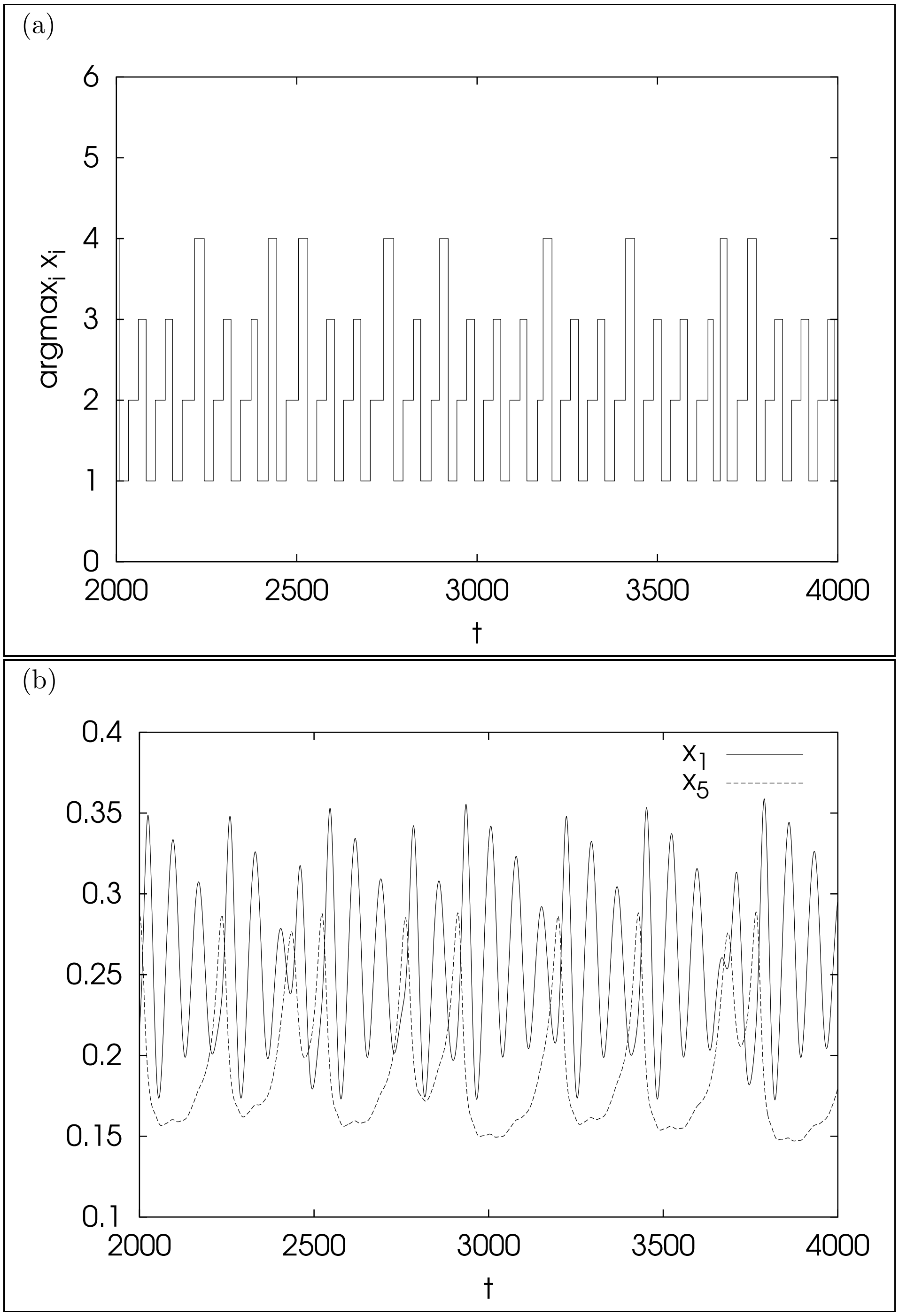


Figure 3.2.3. More about the chaotic orbit in Figure 3.2.2d: **(a)** Dominance plot; the height at time t is the index of the most populous grammar $= \text{argmax}_i\, x_i(t)$. **(b)** Time trace, with x_1 and x_5 plotted against time.

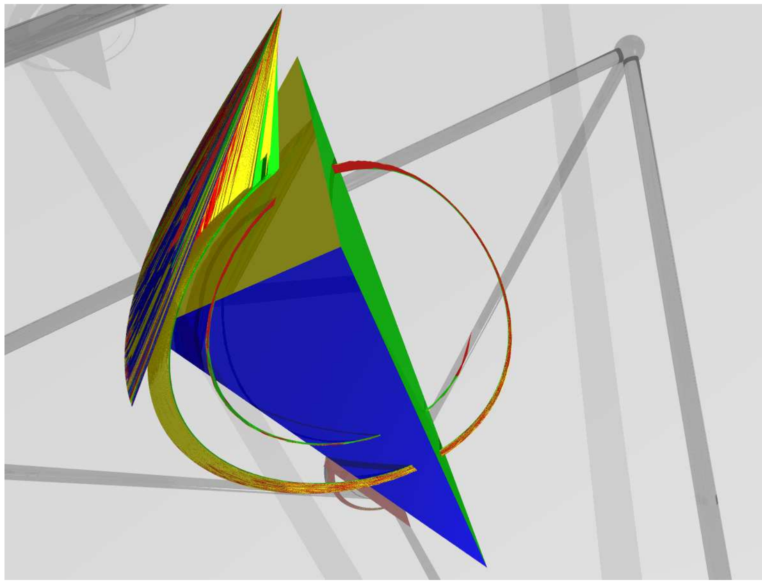


Figure 3.3.1. A Poincaré map for $\mu = 0.73$. The domain of the map is the set of points in the simplex with $x_4 = 0.19$. This picture shows how the map transforms the subset of that domain contained in the multi-colored pyramid. The return image, that is, the result of flowing the pyramid forward through the vector field until it returns to $x_4 = 0.19$, is the sail shape with the spiral attached. Although this picture is not conclusive, it suggests that part of the domain maps back to itself in such a way as to create a generalization of Smale's horseshoe, resulting in chaotic behavior. The glassy rods are the outline of a three-dimensional reference simplex surrounding the domain of the Poincaré map. The mirror plane underneath shows shadows and reflections of the reference simplex, the multi-colored pyramid, and the return image.

picture where it will be dominated by G_5, G_6, and G_7, until the second escape mechanism kicks in, moving it via G_8 back to the left. While the population is on the right-hand side of the picture, features common to G_5, G_6, and G_7 will linger. Thus, it is possible to use the language dynamical equation to describe changes of modules, or sets of features, on different time scales. Populations can spend a long time dominated by a cluster of grammars with a common module. During this time there are rapid changes among the grammars within the cluster. Eventually the populations leaves the cluster and moves to a different cluster with a different module. An arbitrary number of spirals can be combined in this way, yielding extremely complex behavior in any number of dimensions.

3.5. Conclusion

The regular and chaotic oscillations displayed here capture two important features observed in actual languages. First, languages often change spontaneously, following regular patterns such as consonant shifts and changes of morphology type [74]. Thus, for time scales on the order of centuries, the oscillations discussed here are more realistic than the stable equilibria exhibited in Chapter 2. Second, language change is unpredictable and highly sensitive to perturbations. Many language changes, particularly those associated with borrowed

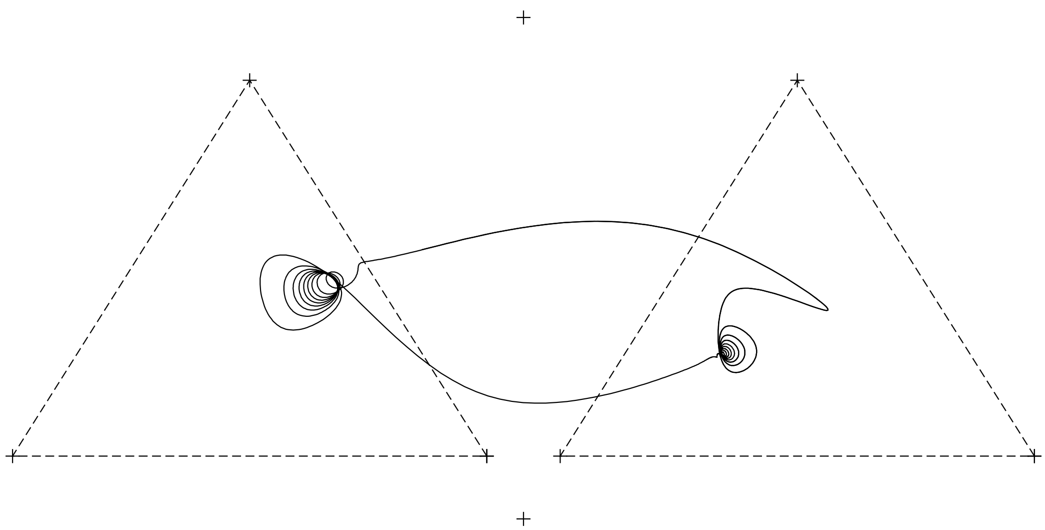


Figure 3.4.1. A complex stable limit cycle among eight grammars, projected from seven dimensions down to two. The triangle on the left represents the face of the phase space spanned by G_1, G_2, and G_3. The triangle on the right represents the face spanned by G_5, G_6, and G_7. The upper cross is the vertex for G_4, and the lower cross is the vertex for G_8. The orbit spirals into the left-hand saddle point, then G_4 starts to increase. Learning errors from G_4 feed into G_5, G_6, and G_7, yielding another spiral. The orbit escapes the spiral around the right-hand saddle point as G_8 increases, and returns to the spiral on the left.

vocabulary, are triggered by language contact. The same kind of unpredictability and sensitivity is exhibited by chaotic dynamical systems.

In summary, the language dynamical equation is a game dynamical equation with learning. Here, we show that complex limit cycles and chaos can arise even for very simple choices of the payoff matrix B and the learning matrix Q. In our example, we considered 5 languages (strategies), each of which is a strict Nash equilibrium. This means that each language is the best reply against itself. A pure replicator dynamics would have 5 stable equilibria corresponding to homogeneous populations where everybody speaks the same language. Then we add a learning matrix, Q, which is diagonally dominant: The most likely outcome of learning is always the correct language. For each target language, however, there is one language which is the second most likely choice. This structure of the learning matrix is sufficient to induce chaos. Thus very conservative, natural choices of payoff matrix and learning matrix lead to deterministic chaos.

Our analysis has implications for historical linguistics, language evolution and evolutionary game theory. Simple learning errors can lead to complex, unpredictable and seemingly stochastic changes in languages over time. For game dynamics, we note that imperfect transmission (learning) of the strategy from one generation to the next can lead to chaotic switching among strict Nash equilibria.

Chapter 4

Competitive Exclusion and Coexistence of Universal Grammars

Contents

†The bulk of this chapter is reprinted from [49]: *Bulletin of Mathematical Biology,* Vol. 65, W. G. Mitchener and M. A. Nowak, Competitive Exclusion and Coexistence of Universal Grammars, Pages 67–93,

4.1. Introduction

While a genetically encoded UG is a logical requirement for the process of language acquisition, there is considerable debate about the nature of the genetically encoded constraints. Interestingly, in a recent study, a mutation in a gene was linked to a language disorder in humans [42] providing a specific example of a genetic modification that affects linguistic performance. It is therefore natural to construct population models which incorporate genetic variation in the form of multiple universal grammars, and to explore the long term behavior of such models.

Universal grammar has no doubt been influenced by natural selection, as well as by mathematical or computational constraints that apply to any communication system, and of course by random chance. It is not clear exactly which aspects of UG are due to natural selection. In particular, many linguists are skeptical of explanations based on proposed adaptive benefits of fine-scale features and particular grammatical rules [45, 76]. However, large-scale properties of UG, such as the number and variety of grammars admitted or the rough form of the learning process, might have broadly acceptable explanations in terms of some notion of adaptive benefit. The purpose of this chapter is to begin exploring this possibility.

Since evolution requires variation, we have to study selection among different UGs, and an obvious place to begin is with an investigation into what happens when more than one UG is present in a given population. As described in Section 1.3, the language dynamical equation may be extended to include multiple UGs, although the number of variables and parameters can easily get out of hand. So, this chapter will focus on cases where one of the UGs admits only one grammar, and discuss what forces favor more specific UGs (those that admit few grammars) or less specific UGs (those that admit many grammars). An interesting finding is that less specific UGs can resist invasion by more specific UGs if learning is more accurate. In other words, accurate learning stabilizes UGs that admit large numbers of candidate grammars.

We explore three possibilities of selective dynamics. The first, *dominance,* means that one particular UG takes over the population from any initial state. The second, *competitive exclusion,* happens when some UG takes over the population, but the initial state influences which one. The third, *coexistence,* means that two or more UGs exist stably. We construct a dynamical system describing a population of individuals. Each individual has an innate UG and speaks one of the grammars generated by this UG. Individuals reproduce in proportion to their ability to communicate with the whole population, passing on their UG to their offspring genetically, and attempting to teach their grammar to their children. The children can make mistakes and learn a different grammar than their parents speak, but within the constraints of their UG.

Section 4.2 briefly reviews the mathematical details of the language dynamical equation with multiple universal grammars, and specifies the particular form of the abstract communication game to be used in this chapter.

Section 4.3 analyzes a one-dimensional case with one UG that specifies two candidate grammars. This simple case is used as a building block for subsequent analysis.

In Section 4.4, we study the selection between two universal grammars: U_1 admits grammar G_1 while U_2 admits grammars G_1 and G_2. This case is of interest because it illustrates the competition between a more specific UG, that is, one with more constraints and therefore fewer options, and a less specific UG. We never find coexistence between U_1 and U_2. For certain parameter values, U_1 dominates U_2, meaning that the only stable equilibrium consists entirely of individuals with U_1. For other parameter values, we find competitive exclusion: Both U_1 and U_2 can give rise to stable equilibria. In particular, U_2 is stable against invasion by U_1 if learning is sufficiently accurate and if most individuals use G_2.

In Section 4.5, we study two extensions. First, we consider what happens if a multi-grammar UG denoted by U_0, which allows grammars G_1 through G_n, competes with n single-grammar UGs denoted by U_1 through U_n, where U_j allows only G_j. To simplify the analysis, symmetry is imposed on the model. It turns out that U_0 is never able to take over the population, but that any one of the single-grammar UGs can. In a second extension, U_0 only competes against U_1. In this case, there can be a stable equilibrium that consists entirely of individuals with U_0, provided its learning algorithm is sufficiently reliable, and the population does not contain too many speakers of G_1.

In Section 4.6 we allow grammars to be ambiguous, and study the case where U_1 admits grammar G_1, while U_2 admits grammars G_2 and G_3. We provide an example where U_2 dominates U_1 and an example where U_1 and U_2 coexist in a stable equilibrium.

In Section 4.7, we draw some conclusions and discuss possible future steps in this line of research.

4.2. Language dynamics with multiple universal grammars

Let us review the multi-UG model from Section 1.3. Suppose we have a large population, each member of which is born with one of the N universal grammars $U_1, U_2, \ldots, U_N$ and speaks one of the n grammars $G_1, G_2, \ldots, G_n$. Each UG consists of a list of which grammars it allows, and has an associated language acquisition algorithm. The grammars are assumed to have an overlap given by the matrix A, where $A_{i,j}$ is the probability that a sentence spoken at random by a speaker of G_i can be parsed by a speaker of G_j. A grammar G_i is said to be *unambiguous* if $A_{i,i} = 1$, because $A_{i,i} < 1$ implies that two people with the same grammar can misunderstand each other due to some sentence with multiple meanings.

Define $x_{j,K}$ to be the fraction of the population which speaks G_j and possesses universal grammar U_K. We have $\sum_K \sum_j x_{i,K} = 1$. Every population state can be represented as a point on a simplex $S_{(Nn)}$. The population changes over time in that individuals reproduce at a rate determined by their ability to communicate with everyone else, passing their universal grammar to their offspring via genetic inheritance, and passing their language on through teaching and learning. We assume that genetic mutation affecting UG is sufficiently rare that such mutations may be treated as isolated events and are not directly included in the equations governing the population dynamics. Instead, mutation will be handled as an external perturbation: Starting from a population where everyone has the same UG, a small fraction is changed to some other UG. Learning errors, in which children mistakenly acquire a grammar different from their parents', are assumed to happen frequently and are modeled directly. The learning process is expressed by the three-axis matrix Q, where $Q_{i,j,K}$ is the probability that a parent speaking G_i produces a child speaking G_j given that both have universal grammar U_K. Since every child must speak some language, Q is row-stochastic, that is, $\sum_j Q_{i,j,K} = 1$ for all i and K. The reproductive rate F_j depends on which grammar an individual uses and the composition of the rest of the population, and is given by

$$F_j = \sum_{K=1}^{N} \sum_{i=1}^{n} B_{i,j} x_{i,K} \quad \text{where} \quad B_{i,j} = \frac{A_{i,j} + A_{j,i}}{2}. \tag{4.2.1}$$

To write the ordinary differential equation (ODE) governing the population dynamics, we also need the variable ϕ which represents the average reproductive rate of the population:

$$\phi = \sum_{K=1}^{N} \sum_{j=1}^{n} F_j x_{j,K}. \tag{4.2.2}$$

The language dynamical equation with multiple universal grammars is then

$$\dot{x}_{j,K} = \sum_{i=1}^{n} F_i x_{i,K} Q_{i,j,K} - \phi x_{j,K} \text{ where } j = 1 \ldots n, K = 1 \ldots N. \tag{4.2.3}$$

In some cases, such as the one in Section 4.4, we will further restrict our attention to a face of $S_{(Nn)}$, which is itself a lower-dimensional simplex. This restriction comes from assuming that some U_K disallows some G_j, so that $x_{j,K}$ is fixed at 0.

4.3. Two grammars and one universal grammar

The case to be examined here, that of a single universal grammar which generates two unambiguous grammars, takes place in S_2, a one-dimensional phase space. We use this case as an essential building block in later sections.

4.3.1. Parameter values. Since there is only one universal grammar, we will omit the K subscript from x and Q. There are three choices of real numbers which fill in all the parameters for this case of the language dynamical equation, which come from considering the possibilities for A and Q as follows. The most general form of the overlap matrix A for two unambiguous grammars is

$$A = \begin{pmatrix} 1 & a_{1,2} \\ a_{2,1} & 1 \end{pmatrix}.$$

However, the A matrix only enters the dynamical system through the B matrix, as in (4.2.1), and since B is a symmetric matrix,

$$B = \frac{A + A^T}{2} = \begin{pmatrix} 1 & (a_{1,2} + a_{2,1})/2 \\ (a_{1,2} + a_{2,1})/2 & 1 \end{pmatrix},$$

there is really only one degree of freedom in choosing A. So, we define

$$b = \frac{a_{1,2} + a_{2,1}}{2}, \tag{4.3.1}$$

and allow this to be the one free parameter determined by the overlap between G_1 and G_2. The most general form for the learning algorithm matrix Q is

$$Q = \begin{pmatrix} q_1 & 1-q_1 \\ 1-q_2 & q_2 \end{pmatrix}, \tag{4.3.2}$$

which has two degrees of freedom. The ranges of the parameters are $0 < b < 1$, $0 < q_1 < 1$, and $0 < q_2 < 1$. Although we can certainly consider the cases where q_1 and q_2 are less than $1/2$, these are somewhat pathological because they represent a situation where children are more likely to learn the grammar opposite to the one their parents speak. Furthermore, if $b = 0$ then G_1 and G_2 have nothing in common and when $b = 1$ they are identical. Both of these settings are degenerate and will not be analyzed here.

4.3.2. Fixed point analysis. In the present case, everything takes place on a unit interval $0 \le x_1 \le 1$, and the dynamical system is one dimensional, as can be seen by expanding (4.2.3) and replacing x_2 with $1 - x_1$:

$$\begin{aligned} \dot{x}_1 = &(1-q_2) \\ &+ (-3 + b(1 + q_1 - q_2) + 2q_2)x_1 \\ &+ (1-b)(3 + q_1 - q_2)x_1^2 \\ &- 2(1-b)x_1^3. \end{aligned} \tag{4.3.3}$$

It is useful to change coordinates to $x_1 = 1 - 2r$ so that the dynamical system inhabits an interval $-1 \le r \le 1$ that is symmetric about 0. The vector field now takes on the form

$$\begin{aligned} \dot{r} = &-\frac{1}{2}\Big((1+b)(q_1 - q_2) \\ &+ (3 + b - 2(q_1 + q_2)))r \\ &+ (1-b)(q_1 - q_2)r^2 \\ &+ (1-b)r^3\Big). \end{aligned} \tag{4.3.4}$$

By straightforward calculation, if $r = -1$ then $\dot{r} = 2(1 - q_1) > 0$, and if $r = 1$ then $\dot{r} = 2(-1 + q_2) < 0$. By the intermediate value theorem, there must be at least one fixed point in the interval. Since $\dot{r}$ is a cubic polynomial in r, there can be either one, two, or three fixed points, depending on the choice of parameters. Keeping in mind that the vector field points inward at both ends of the interval, the dynamical system must follow one of the phase portraits in Figure 4.3.1. Two kinds of bifurcations are possible: saddle-node and pitchfork. The remainder of this section will be spent developing a partial answer to the question of which parameter values cause particular bifurcations, and where the fixed points are when they take place. Rather than solve $\dot{r} = 0$ directly, we will make use of the following variations of some well-known lemmas (see Chapter 1 of [3] or Chapter 4 of [1]) and indirect methods to extract information about the bifurcations.

Lemma 4.3.1. *Let $f(x)$ be a polynomial with a root z of multiplicity $n \ge 1$. Then z is a root of $f'(x)$ with multiplicity $n - 1$.*

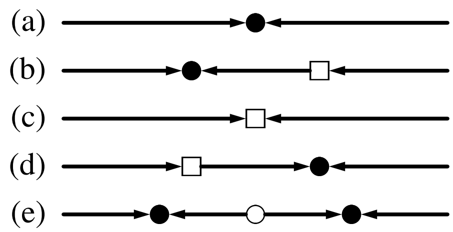


Figure 4.3.1. Possible phase portraits for the base line of the simplex. Key: • indicates a sink, ∘ indicates a source, □ indicates a non-hyperbolic fixed point. Pictures (a) and (e) are structurally stable, (b) and (d) are saddle-node or transcritical bifurcations, and (c) is a pitchfork bifurcation.

Proof. Write $f(x) = (x-z)^n g(x)$ where $g(z) \neq 0$. Then

$$\begin{aligned} f'(x) &= n(x-z)^{n-1}g(x) + (x-z)^n g'(x) \\ &= (x-z)^{n-1}(ng(x) + (x-z)g'(x)). \end{aligned}$$

Observe from the first factor in the bottom line that z is a root of $f'(x)$ of multiplicity at least $n-1$. At $x = z$, the second factor takes the value $ng(z)$ which is nonzero, so the multiplicity of z is exactly $n-1$. □

Lemma 4.3.2. *Let $f(x)$ be a polynomial with a root z such that $f'(z) = 0$. Then z is a root of multiplicity two or more.*

Proof. Let z be a root of f with multiplicity n. Since z is a root of f' of multiplicity $n-1$ and $n-1 \geq 1$, it follows that $n \geq 2$. □

Lemma 4.3.3. *Given a real-valued polynomial dynamical system $\dot{x} = f(x)$, the non-hyperbolic fixed points are exactly the roots of f of multiplicity two or more.*

Proof. From Lemma 4.3.1, every root of f of multiplicity two or more is a non-hyperbolic fixed point. Conversely, if z is a non-hyperbolic fixed point, then $f(z) = 0$ and $f'(z) = 0$, and Lemma 4.3.2 guarantees that z is a root of f of multiplicity two or more. □

Lemma 4.3.3 is the most useful, as it allows us to find the bifurcation parameters of (4.3.4) without explicitly solving a cubic. In particular, for saddle-node and transcritical bifurcations there is a double root of the polynomial and for pitchfork bifurcations there is a triple root of the polynomial. Thus, the parameter settings which generate the non-hyperbolic fixed points in Figure 4.3.1 parts (b), (c), and (d) may be found by matching (4.3.4) against a general template polynomial with multiple roots, as will be illustrated below.

As a side note, the results of this section will be used to analyze higher dimensional dynamical systems in which both saddle-node and transcritical bifurcations will be possible, both of which are characterized by a double root. Saddle-node bifurcations are distinguished from transcritical bifurcations in that the double root comes into existence at the bifurcation rather than forming from the collision of two pre-existing fixed points. The template polynomial method does not distinguish between these two cases as it can only locate parameter settings that produce non-hyperbolic fixed points. The way in which the parameters change so as to pass through such settings determines which type of bifurcation takes place.

Proposition 4.3.4. *The unique parameter setting which produces the phase portrait given in Figure 4.3.1 (c) (the pitchfork bifurcation) is*

$$q_1 = q_2 = \frac{3+b}{4}.$$

The non-hyperbolic fixed point is at $r = 0$, corresponding to $x_1 = x_2 = 1/2$, the center of the phase space.

Proof. The technique is to set $\dot{r} = 0$ and seek parameters that generate a triple root. We divide the resulting cubic equation by the coefficient of r^3 to produce a monic polynomial, and set the resulting coefficients equal to the corresponding coefficients of $(r-p)^3$ where p is an unknown variable, corresponding to the non-hyperbolic fixed point. The resulting system of equations is

$$-p^3 = \frac{(1+b)(q_1 - q_2)}{1-b}, \tag{4.3.5a}$$

$$3p^2 = \frac{3+b-2q_1-2q_2}{1-b}, \tag{4.3.5b}$$

$$-3p = q_1 - q_2. \tag{4.3.5c}$$

It turns out that this system can be solved for q_1 and q_2 in terms of b. To begin, we use (4.3.5c) to eliminate q_2 in the (4.3.5a) which yields

$$p^3 + \frac{3(1+b)}{-1+b}p = 0.$$

This equation has three roots,

$$p = 0, \quad p = \pm\sqrt{3}\sqrt{\frac{1+b}{1-b}}.$$

The second and third roots lie outside the interval of interest $-1 \leq p \leq 1$, so the only possible solution is $p = 0$ from which it follows that $q_1 = q_2 = (3+b)/4$. □

The cases in which there are two fixed points and one is a double root is significantly more complicated because there is an additional unknown variable. This next result is a partial solution.

Proposition 4.3.5. *For the phase portraits shown in Figure 4.3.1 parts (b) and (d) (which are saddle-node or transcritical bifurcations), the sink and non-hyperbolic fixed point lie on opposite halves of phase space.*

Proof. We begin as in Proposition 4.3.4, but this time matching $\dot{r} = 0$ against the cubic template $(r-p_1)^2(r-p_2)$ where p_1 is the non-hyperbolic fixed point and p_2 is the sink.

Since both fixed points are assumed to exist, it must be true that $|p_1| \leq 1$ and $|p_2| \leq 1$. The initial system of equations is

$$-p_1^2 p_2 = \frac{(1+b)(q_1 - q_2)}{1-b}, \tag{4.3.6a}$$

$$p_1^2 + 2p_1p_2 = \frac{3+b-2q_1-2q_2}{(1-b)}, \tag{4.3.6b}$$

$$-2p_1 - p_2 = q_1 - q_2. \tag{4.3.6c}$$

We proceed by solving for p_1 in terms of p_2. Substituting (4.3.6c) into (4.3.6a) results in a quadratic equation in p_1,

$$\left(\frac{1+b}{1-b}\right)(2p_1 + p_2) = p_1^2 p_2,$$

whose roots are

$$p_1 = \frac{C}{p_2} \pm \sqrt{\frac{C^2}{p_2^2} + C} \quad \text{where} \quad C = \frac{1+b}{1-b} > 1.$$

From here, we demonstrate that $p_2 > 0$ implies $p_1 < 0$. Clearly

$$\sqrt{\frac{C^2}{p_2^2} + C} > 1,$$

which implies that the $+$ root lies outside the phase space and is therefore extraneous. Hence the non-hyperbolic fixed point must be located at the $-$ root. It is easy to see that

$$\sqrt{\frac{C^2}{p_2^2} + C} > \frac{C}{p_2},$$

from which it follows that

$$p_1 = \frac{C}{p_2} - \sqrt{\frac{C^2}{p_2^2} + C} < 0.$$

A similar argument shows that $p_2 < 0$ implies $p_1 > 0$. If $p_1 = p_2 = 0$, we have the case of Proposition 4.3.4 which is a different phase portrait. □

This next proposition is a constraint that is needed in Section 4.4.

Proposition 4.3.6. *There is no setting of the parameters for which three fixed points lie on the same side of the middle.*

Proof. Suppose that we start at parameter values for which there is only one fixed point, and change them smoothly so that there are three afterward. This means the system must undergo either a saddle-node or pitchfork bifurcation. In the case of a saddle-node bifurcation, Proposition 4.3.5 ensures that the two new fixed points lie on the other side of the middle from the original fixed point. If a pitchfork bifurcation happens, it must occur at the middle of the phase space according to Proposition 4.3.4, and the two new fixed points must lie to either side of it.

Now assume that three fixed points do exist, and the parameters change so that one of them crosses the middle, that is, at $r = 0$, we have $\dot{r} = 0$. Plugging this assumption into the dynamical system in (4.3.4) implies that $q_1 = q_2$. Thus in this circumstance,

$$\dot{r}|_{q_2=q_1} = \frac{1}{2}r(4q_1 - 3 - b - (1-b)r^2),$$

so the other two fixed points must be at

$$\pm\sqrt{\frac{4q_1 - 3 - b}{1 - b}}.$$

Therefore, the only fixed point which can cross the middle of the phase plane is the central one. □

The complete set of bifurcation parameters can be found implicitly by building from Lemma 4.3.3 and using the discriminant. By definition, the discriminant of a polynomial is the product of the squares of the pair-wise differences of its roots, so it will be zero when a polynomial has a multiple root. The discriminant can be expressed entirely in terms of the coefficients of the polynomial. For a general cubic $a_3z^3 + a_2z^2 + a_1z + a_0$, the discriminant is

$$\frac{a_1^2a_2^2 - 4a_0a_2^3 - 4a_1^3a_3 + 18a_0a_1a_2a_3 - 27a_0^2a_3^2}{a_3^4}. \tag{4.3.7}$$

For $\dot{r}$, the discriminant is a large expression in terms of q_1, q_2, and b obtained by filling in this general formula. The bifurcation parameters are the values of q_1, q_2, and b which make this expression zero, and that surface may be plotted implicitly, as shown in Figure 4.3.2. According to the picture, the surface consists of two curved surfaces which meet in a spine where $q_1 = q_2 = (3+b)/4$ (the pitchfork bifurcation). The bottom corner is at $q_1 = q_2 = 3/4$, $b = 0$, and the rest of the surface appears to lie in the region $q_1, q_2 > (3+b)/4$. The important thing to notice is that if q_1 and q_2 are both close to 1, that is, under the surface, then the dynamical system has three hyperbolic fixed points. Above the surface, there is a single hyperbolic fixed point, and on the surface, there are one or two fixed points with at least one non-hyperbolic. The closer b is to 1, the larger q_1 and q_2 must be to be under the surface.

4.4. Two grammars and two universal grammars

In this section, we analyze a two-dimensional, asymmetric instance of the language dynamical equation. It models the following scenario: Suppose the population has a universal grammar U_1 which generates exactly one grammar G_1; learning and communication are both perfect. Under what circumstances could the population shift in favor of a new universal grammar U_2 which generates G_1 plus an additional grammar G_2? That is, within this model, when is it advantageous to have a choice between two grammars? The analysis builds heavily on the results from Section 4.3.

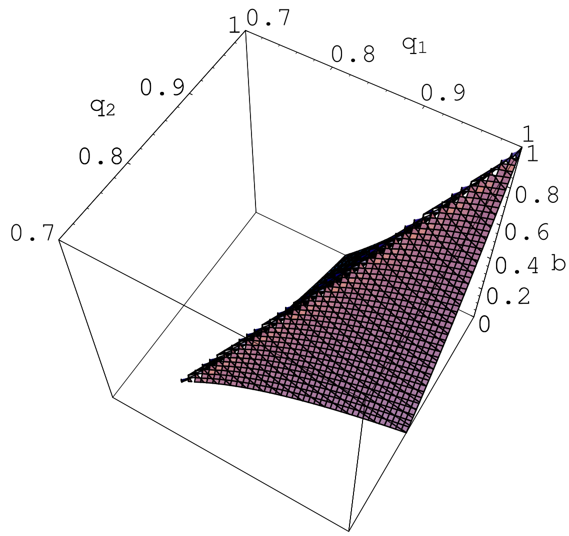


Figure 4.3.2. Bifurcation surface. Observe that in this picture, the q_1 and q_2 axes run only from 0.7 to 1. On the tent-shaped surface, there are one or two fixed points with at least one non-hyperbolic. Above the surface, the system has one hyperbolic fixed point, and below, it has three.

4.4.1. Parameter settings. The dependent variables of interest are $x_{1,1}$, $x_{1,2}$, and $x_{2,2}$. The variable $x_{2,1}$ represents the part of the population which speaks G_2 but has universal grammar U_1, and by assumption, this is zero. Thus, the dynamical system in this case is in three variables with two degrees of freedom and can therefore be analyzed as a planar system.

As in Section 4.3, the A matrix only enters the dynamical system through the B matrix, as in (4.2.1), and since B is a symmetric matrix, there is really only one degree of freedom in choosing A. So, we define

$$b = \frac{a_{1,2} + a_{2,1}}{2}, \tag{4.4.1}$$

and allow this to be the one free parameter determined by the overlap between G_1 and G_2. The most general form for the learning algorithm matrix Q is

$$Q_{i,j,1} = \begin{pmatrix} 1 & 0 \\ * & * \end{pmatrix}, \qquad Q_{i,j,2} = \begin{pmatrix} q_1 & 1 - q_1 \\ 1 - q_2 & q_2 \end{pmatrix}, \tag{4.4.2}$$

which has two degrees of freedom. The entries filled with $*$ are always multiplied by $x_{2,1}$ which is assumed to be 0, so they do not matter. Thus, this model has a total of three free parameters: b, q_1, and q_2, all of which are assumed to lie strictly between 0 and 1.

4.4.2. Geometric analysis of the dynamics. With these parameter settings, and the fact that $x_{1,2} = 1 - x_{1,1} - x_{2,2}$, the dynamical system (4.2.3) simplifies to

$$\begin{aligned}\dot{x}_{1,1} =& -(1-b)x_{1,1}x_{2,2}(2x_{2,2}-1),\\ \dot{x}_{2,2} =& 1 - q_1 + (-1+q_1)x_{1,1}\\ &+ (-3+b+q_1+(1-b)q_1+bq_2+(-1+b)(-1+q_1)x_{1,1})x_{2,2}\\ &+ (-2(-1+b)+(-1+b)(-1+q_1)+q_2-bq_2)x_{2,2}^2\\ &+ 2(-1+b)x_{2,2}^3.\end{aligned} \tag{4.4.3}$$

It lives on the three-vertex simplex S_3, that is, a triangle. The vertices correspond to $x_{j,K} = 1$ and will be labeled $X_{j,K}$ in diagrams.

From here, a fairly complete understanding of the bifurcations of this system can be derived from some simple calculations and geometric considerations. To begin, we will find lines along which $\dot{x}_{1,1} = 0$, and the vector field is therefore parallel to the base of the simplex. These are called $x_{1,1}$ *null-clines*. From (4.4.3), it is clear that $\dot{x}_{1,1}$ is zero in three places: the lines $x_{1,1} = 0$, which is the base of the simplex, and $x_{2,2} = 0$, which is the left edge, and the line $x_{2,2} = 1/2$, which runs across the simplex. In particular, the base line $x_{1,1} = 0$ is invariant under this vector field. See Figure 4.4.1.

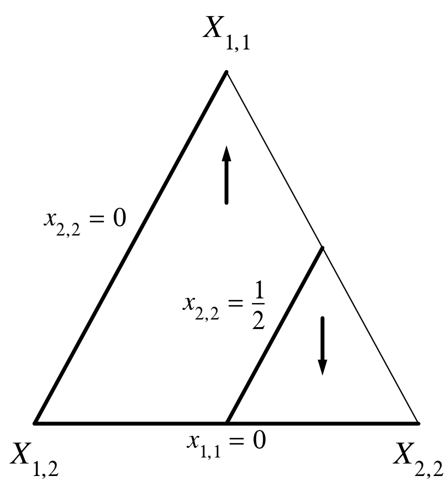


Figure 4.4.1. The simplex, with null-clines. The bold lines indicate where $\dot{x}_{1,1} = 0$. The arrows indicate the sign of $\dot{x}_{1,1}$ in the regions in between, up for positive, down for negative.

Several fixed points are easily located. Observe that if $x_{2,2} = 0$ then $\dot{x}_{1,1} = 0$ and $\dot{x}_{2,2} = (1-q_1)(1-x_{1,1})$. So the apex is the only fixed point on the left side of the simplex. Also, since the vector field always points upward toward it, it is stable. Another fixed point may be located on the cross line by substituting $x_{2,2} = 1/2$ into (4.4.3) yielding

$$\begin{aligned}\dot{x}_{1,1}|_{x_{2,2}=1/2} &= 0,\\ \dot{x}_{2,2}|_{x_{2,2}=1/2} &= \frac{1}{4}(1+b)\left(q_2 - q_1 - 2(1-q_1)x_{1,1}\right),\end{aligned} \tag{4.4.4}$$

from which we find that

$$(x_{1,1}, x_{2,2}) = \left(\frac{q_2 - q_1}{2(1-q_1)}, \frac{1}{2} \right)$$

is the unique fixed point on the line $x_{2,2} = 1/2$. It is located inside the simplex for $q_2 \geq q_1$ and outside otherwise. Observe that the vertical component of the vector field is upward above this fixed point, and downward below it, so it must be unstable. The horizontal component of the vector field to its right points leftward, and to its left it points rightward, indicating that locally, orbits flow toward the fixed point from either side. Thus, this fixed point is a saddle.

Consider the base line, which is invariant under this vector field and may therefore be partially analyzed in isolation. It is exactly the same as the general two-grammar problem from Section 4.3, and must look like one of the phase portraits in Figure 4.3.1, except that those pictures show only stability or instability in the horizontal direction. Stability of one of these fixed points in the vertical direction is determined by which side of the cross line it lies on: $\dot{x}_{1,1}$ is positive on the left side, indicating instability, and negative on the right side, indicating stability.

We must determine where the fixed points in Figure 4.3.1 may lie with respect to the point $(x_{1,1}, x_{2,2}) = (0, 1/2)$, which we do by examining the behavior of the saddle point on the cross line $x_{2,2} = 1/2$. The key fact is that the vector field on the cross line points leftward above the saddle point, and rightward below it, and changes direction only at that fixed point. Observe that the vector field at the upper right end of the cross line $(x_{1,1}, x_{2,2}) = (1/2, 1/2)$ is $(\dot{x}_{1,1}, \dot{x}_{2,2}) = (0, -(1/4)(1+b)(1-q_2))$, which points leftward. The direction of the vector field at $(x_{1,1}, x_{2,2}) = (0, 1/2)$ is either left or right, depending on the configuration of fixed points on the base line. If it points to the left, then the fixed point on the cross line must lie outside the simplex because the vector field must point left along the entire segment of the cross line within the simplex. Similarly, if the vector field points to the right at $(0, 1/2)$, then the the saddle point must lie inside the simplex. From previous analysis, the saddle point lies inside the simplex if and only if $q_2 \geq q_1$, so we have a link between the values of q_1 and q_2 and the phase portraits in Figure 4.3.1.

Now we must determine how the saddle point crosses the base line into the simplex. It must pass through the point $(x_{1,1}, x_{2,2}) = (0, 1/2)$. Substituting this point into (4.4.4), we see that the parameter values which cause this must satisfy $q_1 = q_2$. As it crosses the base line, it must coincide exactly with one of the fixed points there. Since the saddle point passes through the collision, the fixed points must cross in a transcritical bifurcation. To determine which fixed point is crossed, we substitute $q_2 = q_1 = q$ into the dynamical system in (4.4.3) and examine the base line. Note that $x_{1,1} = 0$ so $\dot{x}_{1,1} = 0$. Also:

$$\dot{x}_{2,2}\big|_{q_2=q_1=q, x_{1,1}=0} = (-1 + 2x_{2,2}) \left(-1 + q + (1-b)x_{2,2} + (-1+b)x_{2,2}^2 \right). \tag{4.4.5}$$

The roots of this cubic correspond to the fixed points on the base line; they are

$$\frac{1}{2} \quad \text{and} \quad \frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{4q-3-b}{1-b}}. \tag{4.4.6}$$

We now get three cases. If $q > (3+b)/4$, then there are three fixed points as in Figure 4.3.1 (e), one exactly in the middle and two to either side. If $q < (3+b)/4$, then there is one fixed point, exactly in the middle as in Figure 4.3.1 (a). If $q = (3 + b)/4$, then there is one degenerate fixed point exactly in the middle as in Figure 4.3.1 (c), in which case the pitchfork and transcritical bifurcations happen simultaneously. At any rate, the saddle point can only enter the simplex by passing through the central fixed point on the base line.

The parameter space breaks up into four regions as shown in Figure 4.4.2. The tent-shaped surface is the same as the one in Figure 4.3.2. For parameter settings above it, there is one fixed point on the base line. For parameter settings below it, there are three fixed points on the base line, two on one side of the middle and one on the other. On the faces, there are two fixed points, one non-hyperbolic, and on the edge, there is one non-hyperbolic fixed point. The vertical plane separates the regions where $q_1 < q_2$ from the regions where $q_2 < q_1$. The complete bifurcation scenario is shown in Figure 4.4.3. The fixed points on the base line are constrained by Propositions 4.3.4, 4.3.5, and 4.3.6, so the cases shown are the only possibilities. Phase portraits in Figure 4.4.3 are labeled according to which part of the parameter space in Figure 4.4.2 they represent.

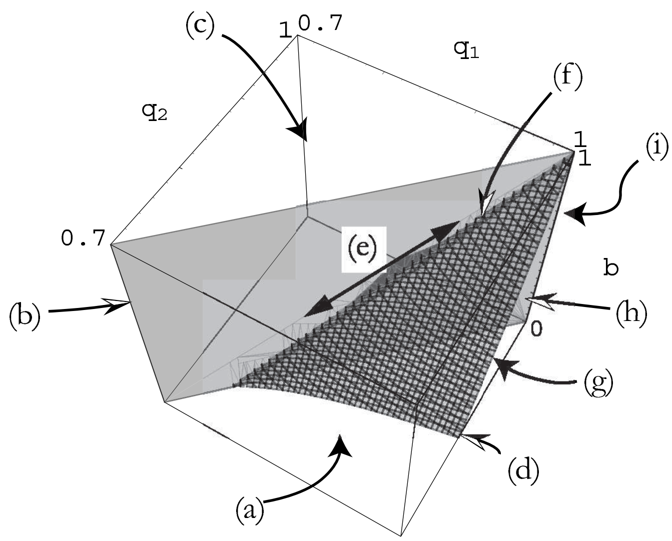


Figure 4.4.2. Parameter space with bifurcation surfaces. Areas (a), (c), (g) and (i) are regions in space, indicated by bold arrows. Areas (b), (d), (f), and (h) are the surfaces that separate those regions, indicated by black and white arrows. Areas (d) and (f) are the front and back surfaces of the tent. Area (b) is the part of the plane above the tent, and (h) is the part below it. Area (e) is a line where the two curved surfaces and the vertical plane intersect, indicated by thin arrows.

4.4.3. Competition between the universal grammars. The bifurcation scenario depicted in Figure 4.4.3 can be analyzed in terms of competition between the two universal grammars. The structurally stable pictures are (a), (c), (g), and (i); these are the ones that occur generically. Observe that in (a), there is only one stable fixed point, and it occurs at the apex of the triangular phase space. All interior orbits will approach this fixed point. Thus, in the case where $q_2 < q_1$ and both are fairly small, U_1 dominates. In (c), there are two stable fixed points, the one at the apex corresponding to a takeover by U_1 and the one

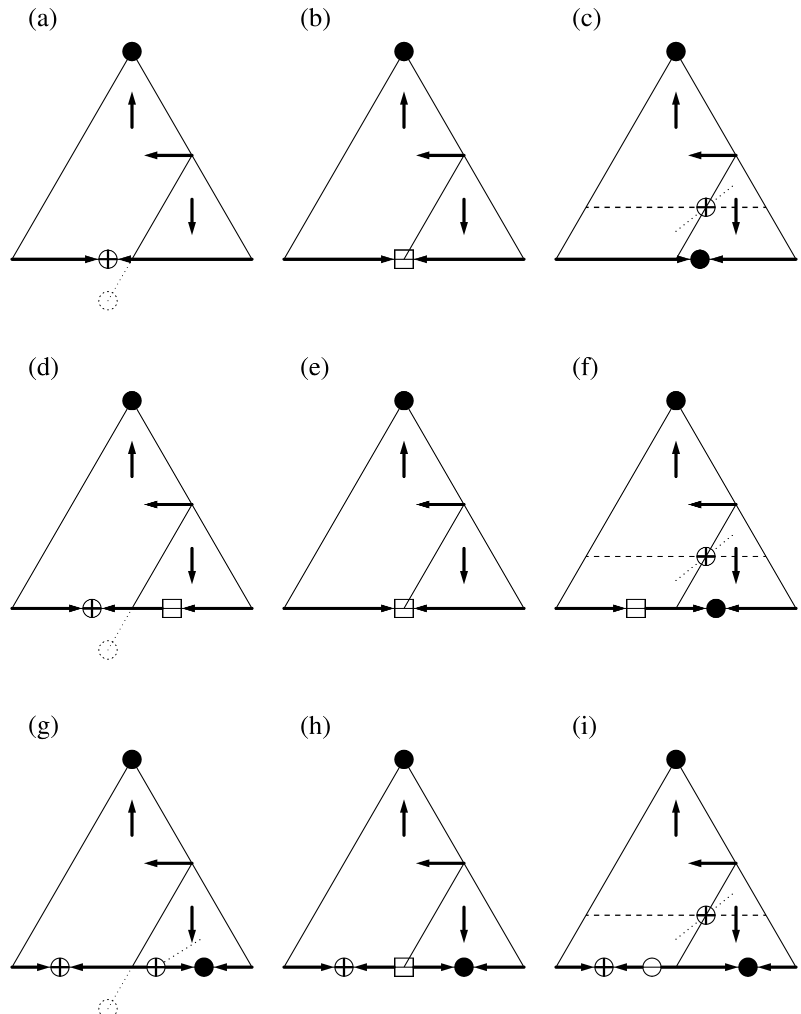


Figure 4.4.3. Phase portraits for the selection dynamics between U_1 (apex) and U_2 (base line). U_1 admits G_1, and U_2 admits G_1 and G_2. Either U_1 dominates as in (a), or there is bistability between U_1 and U_2. The parameters for each picture come from the region of the same label in Figure 4.4.2. Key: $\bullet$ indicates a sink, $\oplus$ indicates a saddle, $\circ$ indicates a source, $\square$ indicates a non-hyperbolic fixed point. Arrows indicate (roughly) the direction of the vector field. In pictures (c), (f), and (i), the cross line and the horizontal dashed line through the saddle point define approximate upper and lower trapping regions for the two sinks. The actual boundary between their basins of attraction is the stable manifold of the saddle point, which is sketched as a dotted line. Picture (g) also contains such a boundary.

on the base line corresponding to a takeover by U_2. Their basins of attraction are separated by the stable manifold of the saddle point on the cross line. Approximations to their basins of attraction can be found by drawing a dashed horizontal line through the saddle. Orbits can only cross the left-hand segment of the dashed line by going upward, and the upper segment of the cross line by going leftward, which means these two segments bound a trapping region containing the apex. Similarly, orbits can only cross the right-hand segment of the dashed line by going downward, and the lower segment of the cross line by going rightward, which means there is another trapping region containing the sink on the base line. In this case, where $q_1 < q_2$ and both are fairly small, there is competitive exclusion between the two universal grammars. The most direct transition from (a) to (c) is a transcritical bifurcation passing through (b). Shortly after this bifurcation, the saddle point will be very close to the base line, so the trapping region for U_2 will be quite small. As q_2 increases, the saddle point moves upward and trapping region expands. A similar situation exists in (i), the difference being that the base line contains two other fixed points which affect a negligible fraction of the phase space. The situation is slightly different in (g). Again, there are two stable fixed points, but the saddle point whose stable manifold separates their basins of attraction is on the base line rather than on the cross line. There does not seem to be a simple trapping region that approximates the basins of attraction in this picture.

4.4.4. Discussion of Section 4.4. To summarize, the scenario examined in this section generically contains instances where U_1 dominates, and instances where there is competitive exclusion, but none where U_2 dominates or where both universal grammars coexist. Furthermore, U_2 can only take over if $q_2 > q_1$ as in pictures (c) and (i), or if q_1 and q_2 are both close to 1 as in picture (g). In the first case, G_2 is acquired more accurately than G_1, so it has an advantage and tends to increase in the population thereby putting U_1 at a disadvantage. In the second case, it appears that although G_1 may be learned more reliably than G_2, the learning reliability of G_2 is sufficiently high that it can maintain a large portion of the population through "market share" effects, again putting U_1 at a disadvantage. Observe that in any case, U_2 can only take over the population through G_2. A population of U_2 people speaking G_1 can be invaded by U_1. This is an illustration of a process by which a valuable acquired trait can become innate.[1] This effect suggests that human universal grammar may have once allowed many more possible grammars than it does now, and that as portions of popular grammars became innate, UG became more restrictive.

4.5. A multi-grammar UG competing with single-grammar UGs

In this section, we will examine cases in which a UG with multiple grammars competes with a number of UGs that have only a single grammar each. We will begin by building on the results from Section 4.4 in two ways, extending that analysis to symmetric cases in an arbitrary number of dimensions.

4.5.1. The case of full competition. Let us extend the case from Section 4.4 by assuming that there are three universal grammars. The first, U_1, allows only G_1. The second, U_2, allows only G_2. The third, U_0, allows both G_1 and G_2. Since there is one single-grammar UG for each possible grammar, this case will be called *full competition.* We would like to determine whether one of these UGs can take over the population.

[1] Also, see Section sec:cx:sca-discussion for a discussion of accidental stability.

This situation contains two copies of the case from Section 4.4, one in which everyone uses U_0 or U_1, and a second in which everyone uses U_0 or U_2. From the former, there is generically no stable equilibrium in which U_0 takes over with a majority of people speaking G_1. From the latter, there is generically no stable equilibrium in which U_0 takes over with a majority of people speaking G_2. If U_0 is to take over, either G_1 or G_2 must be in the majority, so it follows that U_0 is unable to take over.

This result extends to an arbitrary number of grammars as follows. Let the grammars be G_1 to G_n, and assume there are universal grammars U_i which specify only the grammar G_i. Assume there is an additional UG U_0 which allows any of the n grammars. As a simplification, assume that the grammars are fully symmetric and unambiguous, that is, $A_{i,i} = 1$ and $A_{i,j} = a$ for $i \neq j$. The parameter a is required to be strictly between 0 and 1. For reasons that will become clear in a moment, the learning matrix Q is allowed to be fully general except that no grammar is allowed to have perfect learning under U_0, that is, $Q_{i,i,0} < 1$ for all i.

We will need the following new notation. We are interested in determining if one universal grammar out of the U_K can take over the population, and if so, which one. We therefore define

$$y_K = \sum_{j=1}^{n} x_{j,K} \tag{4.5.1}$$

to be the total population with U_K. The dynamics for y_K can be expressed succinctly by using the fact that Q is row stochastic:

$$\begin{aligned}
\dot{y}_K &= \sum_{j=1}^{n} \dot{x}_{j,K} \\
&= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} (F_i x_{i,K} Q_{i,j,K}) - \phi x_{j,K} \right) \\
&= \sum_{i=1}^{n} \left(F_i x_{i,K} \sum_{j=1}^{n} Q_{i,j,K} \right) - \phi \sum_{j=1}^{n} x_{j,K} \\
&= \sum_{i=1}^{n} F_i x_{i,K} - \phi y_K.
\end{aligned} \tag{4.5.2}$$

We may further simplify the notation by introducing the variables

$$\phi_K = \sum_{i=1}^{n} F_i x_{i,K}, \tag{4.5.3}$$

from which it follows that $\phi = \sum_K \phi_K$ and

$$\dot{y}_K = \sum_{i=1}^{n} F_i x_{i,K} - \phi y_K = \phi_K - \phi y_K. \tag{4.5.4}$$

There is no explicit reference to Q in $\dot{y}_K$, although Q does influence the dynamics. It happens that the main result of this section does not depend on Q for exactly this reason.

Because of the symmetry imposed on A, the dynamics of the y_K simplify considerably. If we further define

$$v = (x_{1,0}, x_{2,0}, \dots, x_{n,0}), \tag{4.5.5}$$

$$w = (x_{1,1}, x_{2,2}, \dots, x_{n,n}), \tag{4.5.6}$$

then

$$\begin{aligned} \dot{y}_0 &= -(1-a)((v+w)\cdot w)y_0, \\ \dot{y}_K &= (1-a)(x_{K,0} + y_K - (v+w)\cdot(v+w))y_K \text{ where } K = 1\dots n. \end{aligned} \tag{4.5.7}$$

Note that the sum of the entries of v is equal to y_0, and the sum of the entries of w is equal to $1 - y_0$.

Proposition 4.5.1. *The multi-grammar universal grammar, U_0, is always unstable, that is, if $y_0 < 1$, then $\dot{y}_0 < 0$. The single-grammar UGs are stable, meaning that for $K \geq 1$, if y_K is close to 1, then y_K is increasing.*

Proof. We will prove both statements by starting from a population that consists entirely of one UG, and perturbing it by converting ε of the population to another UG.

To prove the first statement, suppose that $y_0 = 1 - \varepsilon$. All the entries of v and w are greater than or equal to zero, so $v \cdot w \geq 0$. Since w must be non-zero, it follows that $\dot{y}_0 = -(1-a)(v \cdot w + w \cdot w)y_0 < 0$. In fact, in any population state where not everyone has U_0, the number of people with U_0 will decrease. Thus, U_0 is unstable and cannot take over the population.

To prove the second statement, fix $K \geq 1$ and assume that $y_K = 1 - \varepsilon$. Observe that

$$\begin{aligned} (v+w)\cdot(v+w) &= \sum_{i=1}^{n} (x_{i,0} + x_{i,i})^2 \\ &= (x_{K,0} + 1 - \varepsilon)^2 + \sum_{i \neq K} (x_{i,0} + x_{i,i})^2. \end{aligned}$$

The summation is over $n-1$ terms, each of which greater than or equal to zero, and their sum is fixed at $1 - (1-\varepsilon) - x_{0,K}$. Therefore, the summation is at most $(\varepsilon - x_{0,K})^2$. (See Lemma 4.5.2.) It follows that

$$\begin{aligned} \dot{y}_K &\geq (1-a)(1-\varepsilon)\left(1 - \varepsilon + x_{K,0} - (1 - \varepsilon + x_{K,0})^2 - (\varepsilon - x_{K,0})^2\right) \\ &= (1-a)(1-\varepsilon)(\varepsilon - x_{K,0})(1 - 2(\varepsilon - x_{K,0})). \end{aligned}$$

As long as $x_{K,0} < \varepsilon$, we have $\dot{y}_K > 0$. This will continue to be true as $x_{K,0} \leq y_K$.

If $x_{K,0} = \varepsilon$, that is, $x_{K,0}$ accounts for the entire perturbation, then we need the assumption that under U_0, no language is learned perfectly. So, a short time later, $x_{K,0}$ will decrease as some children will have mistakenly learned another grammar, say G_h, so $x_{h,0} > 0$. At this point, we will have a new perturbation with $y_K = \varepsilon'$ and $x_{K,0} < \varepsilon'$, and it follows that $\dot{y}_K > 0$. □

The following lemma is used to make approximations in this and other proofs in this chapter.

Lemma 4.5.2. *Suppose that for $i = 1 \dots m$, we have numbers $\alpha_i \geq 0$ such that $\sum_i \alpha_i = \sigma$. Then*

$$\frac{\sigma^2}{m} \leq \sum_{i=1}^{m} \alpha_i^2 \leq \sigma^2.$$

Proof. Consider $\alpha = (\alpha_i)_{i=1}^m$ as a vector in $\mathbf{R}^m$. It is contained in a simplex because the sum of its entries is fixed. The point on the simplex closest to the origin is the center, corresponding to $\alpha_i = \sigma/m$ for all i, and this point yields the lower bound. The vertices of the simplex are the farthest points from the origin, corresponding to $\alpha_j = \sigma$ and $\alpha_i = 0$ for all $i \neq j$, and these points give the upper bound. $\square$

Proposition 4.5.1 implies that UGs with many grammars are unable to compete directly with UGs that specify only one grammar.

4.5.2. The case of limited competition. The two-dimensional case from Section 4.4 illustrates a situation where a multi-grammar UG can have a stable equilibrium where a majority of the people use a grammar that does not occur as part of a single-grammar UG. We now turn our attention to a different extension of this case in which there are two UGs, U_0 which specifies $G_1, \dots, G_n$, and U_1 which specifies only G_1. As before, the A matrix is assumed to be fully symmetric, with all diagonal entries $A_{i,i} = 1$ and all off-diagonal entries $A_{i,j} = a$. The Q matrix disappears again, and we need only the assumption that no grammar is learned perfectly under U_0. By using the fact that $y_1 = x_{1,1} = 1 - y_0$, the model can be reduced to one differential equation of interest,

$$\begin{aligned} \dot{y}_0 &= (1-a)(-x_{1,1} - x_{1,0} + 2x_{1,1}x_{1,0} + M_2)x_{1,1} \\ &= (1-a)(-1 + y_0 - x_{1,0} + 2(1-y_0)x_{1,0} + M_2)(1-y_0). \end{aligned} \tag{4.5.8}$$

where

$$M_k = \sum_{j=1}^{n} \sum_{K=1}^{N} x_{j,K}^k.$$

There is a fixed point at $y_0 = 0$, as can be seen by substituting this state into the differential equation. Furthermore, $\dot{y}_0 = 0$ when $y_0 = 1$, so the model can have trapping regions and stable fixed points in the subset of states which satisfy $y_0 = 1$. We are interested in determining when these various states are stable under perturbations.

Proposition 4.5.3. *The fixed point $y_0 = 0$, corresponding to a takeover by U_1, is stable.*

Proof. Consider a small perturbation, $y_0 = \varepsilon$. Then we must have $y_1 = x_{1,1} = 1 - \varepsilon$, and the differential equation satisfies

$$\begin{aligned} \dot{y}_0 &= (1-a)\left(-1 + \varepsilon - x_{1,0} + 2(1-\varepsilon)x_{1,0} + (1-\varepsilon)^2 + \sum_{j=1}^{n} x_{j,0}^2\right)(1-\varepsilon) \\ &\leq (1-a)(-1 + \varepsilon + x_{1,0}(1-2\varepsilon) + 1 - 2\varepsilon + \varepsilon^2 + \varepsilon^2)(1-\varepsilon), \end{aligned}$$

where we have used Lemma 4.5.2 to bound the summation by ε^2. This expression factors into

$$\dot{y}_0 \le -(1-a)(\varepsilon - x_{1,0})(1-2\varepsilon).$$

If the perturbation is such that $x_{1,0} < \varepsilon$, then the right hand side is negative, and as $x_{1,0} \le y_0$, it will remain negative, so y_0 will shrink to 0.

If the perturbation is such that $x_{1,0} = \varepsilon$, then we must use the fact that under U_0 there is no perfect learning. After a short time, some other part of the population with U_0, say, $x_{h,0}$, will be non-zero due to learning error. This new perturbation will have $y_0 = \varepsilon'$ and $x_{1,0} < \varepsilon'$, and as before y_0 will shrink to 0. ☐

The following results show that U_0 can still take over, but not with G_1. It states that if $x_{1,0}$ is small enough, then a population consisting only of people with U_0 that is perturbed by adding a small number of people with U_1 will recover, at least in the short term.

Proposition 4.5.4. *Let $\varepsilon > 0$ be small and suppose $y_0 = 1-\varepsilon$ and $x_{1,1} = \varepsilon$. Define $\kappa = 1/n - x_{1,0}$. If $\kappa > \varepsilon/(1-2\varepsilon)$, then $\dot{y}_0 > 0$.*

Proof. From the differential equation,

$$\begin{aligned}\dot{y}_0 &= (1-a)\left(-\varepsilon + (2\varepsilon - 1)x_{1,0} + \varepsilon^2 + \sum_{j=1}^{n} x_{j,0}^2\right)\varepsilon \\ &\ge (1-a)\left(-\varepsilon + (2\varepsilon - 1)x_{1,0} + \varepsilon^2 + \frac{(1-\varepsilon)^2}{n}\right)\varepsilon,\end{aligned}$$

where we have once again used Lemma 4.5.2 to bound the summation. By substituting $x_{1,0} = 1/n - \kappa$, the inequality can be simplified to

$$\dot{y}_0 \ge (1-a)\left(\kappa - \varepsilon(1+2\kappa) + \left(1 + \frac{1}{n}\right)\varepsilon^2\right)\varepsilon.$$

The assumption that $\kappa > \varepsilon/(1-2\varepsilon)$ is equivalent to $\kappa > \varepsilon(1+2\kappa)$, so the right hand side is positive. ☐

The tricky part about interpreting this proposition is that a population state with $y_0 = 1$ might sill be unstable in the long term: It could move within the constraint $y_0 = 1$ to a state where $x_{1,0} > 1/n$, at which point Proposition 4.5.4 no longer applies and a perturbation can cause the population to be taken over by U_1, as this next proposition illustrates.

Proposition 4.5.5. *Let $\varepsilon > 0$ be small and suppose $y_0 = 1-\varepsilon$ and $x_{1,1} = \varepsilon$. If $x_{1,0} > 1/2-\varepsilon$, then y_0 is decreasing.*

Proof. For this proof, we use Lemma 4.5.2 to bound the summation in $\dot{y}_0$ from above,

$$\sum_{j=1}^{n} x_{j,0}^2 = x_{1,0}^2 + \sum_{j=2}^{n} x_{j,0}^2 \le x_{1,0}^2 + (1-\varepsilon-x_{1,0})^2.$$

This bound yields the inequality

$$\dot{y}_0 \leq 2\varepsilon(1-a)(x_{1,0}-(1-\varepsilon))\left(x_{1,0}-\left(\frac{1}{2}-\varepsilon\right)\right).$$

If $1-\varepsilon > x_{1,0} > 1/2-\varepsilon$, then $\dot{y}_0$ is negative, and the perturbation will draw the population away from the region where $y_0 = 1$, indicating instability.

If $x_{1,0} = 1-\varepsilon$, then we resort to the argument that a short time later, the population will change due to learning error to a different perturbation where $y_0 = 1-\varepsilon'$ and some other sub-population $x_{h,0} > 0$. Now $x_{1,0} < 1-\varepsilon'$, which implies that $\dot{y}_0 < 0$ and the population is moving away from the region where $y_0 = 1$. □

4.5.3. Some remarks about these results. Several remarks are in order. First, the Q matrix has mostly disappeared, so Propositions 4.5.1, 4.5.3, 4.5.4, and 4.5.5 hold regardless of the learning mechanism under U_0, except that it must not be perfect. In fact, it could be dynamic, depending on the population state for example, as long as it remains row stochastic.

Second, the fact that some of the propositions declare $y_0 = 1$ to be "stable" may be misleading. As noted before, the population could start in a state where $y_0 = 1$ and move within that constraint to a state in which y_0 begins to decrease. The simplest behavior for which $y_0 = 1$ would be truly stable is for the population to converge to a stable fixed point that satisfies Proposition 4.5.4, but it could also converge to a limit cycle or to a strange attractor, depending on what behaviors are available to a population restricted to U_0.

We can get more definite results from these propositions if we add assumptions that ensure that all population states with $y_0 = 1$ tend to fixed points. Any fixed points that are stable when only U_0 is allowed and that also fall under Proposition 4.5.4 are stable with respect to all perturbations, including those involving the introduction of U_1. Any such fixed points that fall under Proposition 4.5.5 are unstable. Some may be outside the hypotheses of both propositions, and we can say nothing more about them here.

A full bifurcation analysis of the fully symmetric case of the language dynamical equation with one universal grammar is worked out in Chapter 2 and [36]. To apply those results here, we must add the assumption that for the learning matrix for U_0, all diagonal elements $Q_{i,i,0} = q$ and all off-diagonal elements $Q_{i,j,0} = (1-q)/(n-1)$. It follows from Proposition 2.7.3 that if only U_0 is present, then all populations tend to fixed points. The analysis shows that there is a constant $\hat{q}_1$ such that if $q < \hat{q}_1$, then the only stable fixed point in a population restricted to U_0 is one in which every grammar is represented equally. Thus, $x_{1,0} = 1/n$ and that fixed point is potentially unstable to perturbations involving U_1 because Proposition 4.5.4 does not apply. On the other hand, if $q > \hat{q}_1$, then there are n stable fixed points, and each G_j is used by a large part of the population in exactly one of them. These are called the *1-up fixed points* in Chapter 2 and *single grammar fixed points* in [36]. At the one where G_1 has the majority, $x_{1,0} > 1/n$, so it does not fall under Proposition 4.5.4 and is potentially unstable to perturbations involving U_1. If q is sufficiently large, this fixed point moves so that $x_{1,0}$ approaches 1, so at some value of q, it will exceed 1/2. Then Proposition 4.5.5 will apply and the fixed point will definitely be unstable. At the other fixed points, $x_{1,0} < 1/n$, so they fall under Proposition 4.5.4 and are therefore stable. In short, if the learning process in U_0 is sufficiently reliable, that is $q > \hat{q}_1$, then U_0 can take over the population in a stable manner, but not through G_1. If learning is unreliable, then U_1 will eventually take over.

4.6. Ambiguous grammars

In this section we will generalize the case in Section 4.4 not by adding dimensions but by allowing the grammar specified by U_1 to be different from both of those specified by U_2, and by allowing the grammars to be ambiguous. The diagonal entries of A are allowed to be less than one. This case can exhibit a greater variety of behavior than was seen in Section 4.4, including stable coexistence of both universal grammars, and dominance by U_2. This form of the language dynamical equation has a total of eight free parameters. Rather than attempt a complete symbolic analysis, we will present one short proposition and some numerical results.

4.6.1. Parameter values. We assume that U_1 allows for one grammar G_1, and that U_2 allows for two grammars, G_2 and G_3. The Q matrix is allowed to be fully general,

$$Q_{i,j,1} = \begin{pmatrix} 1 & 0 & 0 \\ * & * & * \\ * & * & * \end{pmatrix}, \qquad Q_{i,j,2} = \begin{pmatrix} * & * & * \\ 0 & q_2 & 1-q_2 \\ 0 & 1-q_3 & q_3 \end{pmatrix}. \tag{4.6.1}$$

The entries filled with $*$ are always multiplied by some $x_{j,K}$ that is restricted to be zero, so they do not matter. Also, the matrix A is allowed to be fully general; we even allow the diagonal elements to be less than 1. The only constraint we place on A is that since it appears in the model only through $B = (A + A^T)/2$ we may as well assume A is symmetric. There are eight free parameters, six from the upper half of A and q_2 and q_3.

4.6.2. Analysis and phase portraits. The expressions for $\dot{x}_{1,1}$, $\dot{x}_{2,2}$ and $\dot{x}_{3,2}$ are unwieldy so they will not be written out. However, it turns out that $x_{1,1} = 1$, $x_{2,2} = x_{3,2} = 0$ is a fixed point for all parameter settings. The one exact result is the following:

Proposition 4.6.1. *The fixed point $x_{1,1} = 1$, $x_{2,2} = x_{3,2} = 0$ is unstable if $-2A_{1,1} + A_{1,2}q_2 + A_{1,3}q_3 > 0$.*

Proof. We reduce the system to two dimensions by replacing $x_{3,2}$ by $1 - x_{1,1} - x_{2,2}$. The trace of the Jacobian matrix of the reduced system at the fixed point in question is $-2A_{1,1} + A_{1,2}q_2 + A_{1,3}q_3$. If this is positive, then at least one of the eigenvalues of the Jacobian must have positive real part [70, p. 137]. $\square$

Roughly what this proposition means is that if G_1 is sufficiently ambiguous, and G_2 and G_3 are similar to it and can be learned reliably, then U_1 is unable to take over the population. This situation seems unrealistic, however, there is at least one reasonable interpretation. Suppose that G_1 is close to the union of G_2 and G_3, and contains many sentences that can be interpreted so as to have multiple meanings. Suppose further that many of these sentences are in G_2 or G_3 but with a single meaning. Thus, U_2 has an advantage because it restricts its people to some less ambiguous language at the expense of imperfect learning, and this may be enough to destabilize a population where everyone has U_1. Proposition 4.6.1 is a mathematical expression of this situation. Note that when $A_{1,1}$ is restricted to be 1, the proposition never applies, and the stability of the U_1 fixed point must be determined by other means.

A number of phase portraits for a variety of parameter values are drawn in Figures 4.6.1–4.6.4 based on numerical computations. In particular, these phase portraits illustrate that with this general model, it is possible to have stable coexistence of U_1 and U_2, and it is possible for U_2 to dominate. Neither of these situations is possible in the limited case analyzed in Section 4.4.

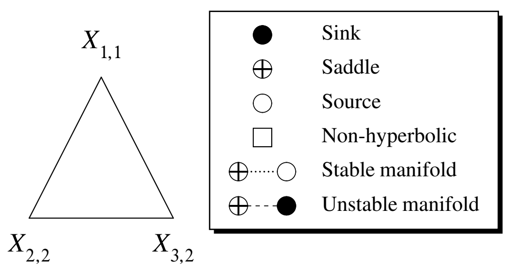
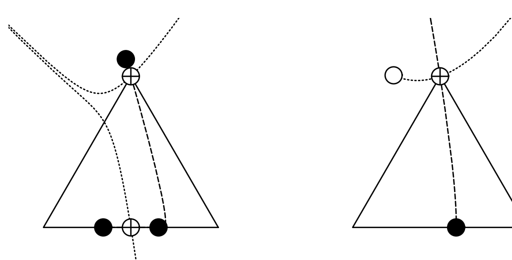


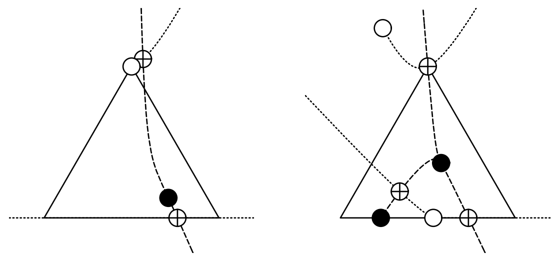
Figure 4.6.1. Key to phase portraits show in Figures 4.6.2 to 4.6.4. Some fixed points outside the simplex have been drawn for reference. The three corners of the triangle represent population states where everyone uses a single language, as indicated. The apex of the triangle represents $U_1 = 1$ and the base represents $U_2 = 1$.



(a) $A = \begin{pmatrix} 0.66 & 0.6 & 0.7 \\ 0.6 & 0.9 & 0.7 \\ 0.7 & 0.7 & 0.9 \end{pmatrix}$, (b) $A = \begin{pmatrix} 0.45 & 0.6 & 0.7 \\ 0.6 & 0.9 & 0.7 \\ 0.7 & 0.7 & 0.9 \end{pmatrix}$,

$q_2 = 0.95, q_3 = 0.95$ $q_2 = 0.8, q_3 = 0.85$

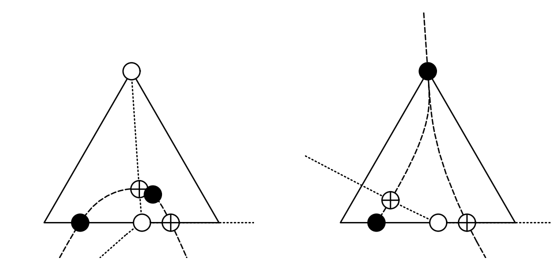
Figure 4.6.2. Two instances where U_2 dominates.



$$\text{(a)}\, A = \begin{pmatrix} 0.4 & 0.6 & 0.7 \\ 0.6 & 0.9 & 0.2 \\ 0.7 & 0.2 & 0.9 \end{pmatrix}, \quad \text{(b)}\, A = \begin{pmatrix} 0.5 & 0.6 & 0.7 \\ 0.6 & 0.9 & 0.2 \\ 0.7 & 0.2 & 0.9 \end{pmatrix},$$

$$q_2 = 0.8, q_3 = 0.85 \qquad q_2 = 0.86, q_3 = 0.85$$

Figure 4.6.3. Two instances of stable coexistence. In (b), U_2 can also take over, but only with G_2.



$$\text{(a)}\, A = \begin{pmatrix} 0.3 & 0.6 & 0.7 \\ 0.6 & 0.9 & 0.2 \\ 0.7 & 0.2 & 0.9 \end{pmatrix}, \quad \text{(b)}\, A = \begin{pmatrix} 0.7 & 0.6 & 0.7 \\ 0.6 & 0.9 & 0.2 \\ 0.7 & 0.2 & 0.9 \end{pmatrix},$$

$$q_2 = 0.87, q_3 = 0.85 \qquad q_2 = 0.87, q_3 = 0.85$$

Figure 4.6.4. Another instance of stable coexistence and an instance of exclusion.

4.7. Conclusion

The evolution of universal grammar is based on genetic modifications that affect the architecture of the brain and the classes of grammars that it can learn. At some point in the evolutionary history of humans, a UG emerged that allowed the acquisition of language with unlimited expressibility. In principle, UG can change as a consequence of random drift (neutral evolution), as a by-product of selection for other cognitive function, or under selection for language acquisition and communication. The third aspect is what we consider in this chapter.

We explore some low-dimensional cases of natural selection among universal grammars. In particular, we study the competition between more specific and less specific UGs. Suppose two universal grammars, U_1 and U_2 are available, and U_2 admits two grammars, G_1

and G_2, while U_1 admits only G_1. If learning within U_2 is too inaccurate, then U_1 dominates U_2: For all initial conditions that include both U_1 and U_2, U_1 will eventually out-compete U_2. If learning within U_2 is sufficiently accurate, then for some initial conditions U_2 will win while for others U_1 will win; there is competitive exclusion. Note that accurate learning stabilizes less specific UGs. We can also find coexistence of two different UGs. We provide such an example where U_1 admits G_1 and U_2 admits G_2 and G_3.

A standard question in ecology is concerned with the competition between specialists that exploit a specific resource and generalists that utilize many different resources [46]. Similarly, here we have analyzed competition between specialist UGs that admit few grammars and generalist UGs that admit many candidate grammars. This is an interesting similarity. There is also a major difference: In ecology the more individuals exploit a resource the less valuable this resource becomes, but in language the more people use the same grammar the more valuable this grammar becomes. Hence, the frequency dependency of the fitness functions work in opposite directions in the two cases.

The question that we ultimately want to understand is the balance between selection for more powerful language learning mechanisms that allow acquisition of larger classes of complex grammars, and selection for more specific UGs that limit the possible grammars. This chapter provides mathematical machinery and a first step toward this end.

Chapter 5

More About Competition Between Universal Grammars

Contents

5.1. Introduction

The intent of this chapter is to further explore the model of how different universal grammars interact and compete for carriers within a population. The analysis in Chapter 4 deals with some two-dimensional cases and addresses some general questions of how the model behaves when one of the UGs in question admits only one grammar. In this chapter, we will address a three-dimensional case of the model, in which there are two UGs, each of which admits two grammars. The goal here is to put aside the issue of selection for more specific or less specific UGs that was addressed in Chapter 4 and focus on the competition between two equally specific UGs. As before three kinds of behavior are of interest: dominance, competitive exclusion, and coexistence.

In Section 5.2, we review the language dynamical equation with multiple universal grammars. From the general case, we restrict our attention to the case of two UGs with two grammars each, and initially impose a simplifying assumption on the set of possible parameters. Section 5.3 contains an almost complete analysis of this case. The assumptions about the parameters are stated in Section 5.3.1. They add a certain amount of symmetry to the dynamical system, and it is very natural to take advantage of this symmetry through a change of coordinates as given in Section 5.3.2. In Section 5.3.3, all fixed points are found algebraically. Many of them lie in one of two invariant planes, and their stabilities are determined in Section 5.3.4. Parameter settings for which the system undergoes a bifurcation within an invariant plane are found as well. Sections 5.3.5 and 5.3.6 assemble these results into complete three-dimensional phase portraits, and include illustrations of dominance, competitive exclusion, and coexistence. The results of this section have interesting implications for the genetic evolution of UG. Section 5.3.7 discusses some of these. In particular, the ability of an individual to communicate with the rest of the population is so important that any mutation introducing a sufficiently incompatible innovation is likely to die out in spite of any other advantage it might carry.

In the second half of this chapter, all restrictions on B are lifted. It turns out that for a range of parameter settings, the two UGs are each stable against invasion by the other. Sufficient conditions on the parameters to imply that homogeneous populations are stable are derived in Section 5.4. In light of the discussion of Section 5.3.7, this situation is important as it describes when a mutated UG is too different from the existing UG to survive long. Section 5.4.1 is a simplified form of the proof in the case of two highly symmetric UGs, and Section 5.4.2 is the complete proofs for general parameter settings. A brief discussion of these results appears in Section 5.4.3.

Finally, in Section 5.5, we draw some conclusions and indicate directions for further research.

5.2. The model

In this chapter, we will use the language dynamical equation with multiple universal grammars, as described in (1.3.5). The variables of primary interest are the fractions of the population with U_K, denoted y_K,

$$y_K = \sum_{j=1}^{n} x_{j,K}. \tag{5.2.1}$$

For ease of notation, we also define variables for the fraction speaking G_j,

$$w_j = \sum_{K=1}^{N} x_{j,K}. \tag{5.2.2}$$

The language dynamical equation for multiple universal grammars is

$$\begin{aligned} F &= Bw, \\ \phi &= w^T Bw, \\ \dot{x}_{j,K} &= \sum_{i=1}^{n} F_i x_{i,K} Q_{i,j,K} - \phi x_{j,K}. \end{aligned} \tag{5.2.3}$$

Section 5.3 assumes two UGs with two grammars each and a particular form for B, and deals with these questions: Under what circumstances are there stable fixed points with $y_K = 1$, thus yielding exclusion or dominance? What about stable fixed points where y_1 and y_2 are both strictly positive, yielding coexistence?

Section 5.4 considers a completely general B matrix. This is a much harder problem, so we focus only on those population states where everyone has the same UG: Suppose $y_K = 1 - \varepsilon$, that is, the population state contains a small invasion. Does it tend back to $y_K = 1$? If so, then the set of all states with $y_K = 1$ forms an attracting set, and U_K is stable against invasion by the other UGs in the model. Thus the dynamics of the y_K variables will be crucial. Since Q is row stochastic, the expression for $\dot{y}_K$ simplifies considerably, leaving just

$$\dot{y}_K = \sum_i F_i x_{i,K} - \phi y_K. \tag{5.2.4}$$

The remarkable fact about 5.2.4 is that Q has disappeared. The learning process still influences the dynamics of y_K in that it steers the $x_{i,K}$. However, as will be shown in Section 5.4, there is the possibility that the overall behavior of the y_K's, and hence the stabilities of U_1 and U_2, can sometimes be determined without reference to Q.

5.3. A three-dimensional case with some symmetry

In this section, we analyze a special case of the dynamical system in three dimensions with enough symmetry that it only has four free parameters. We assume that there are two UGs and four grammars, where U_1 allows G_1 and G_2, and U_2 allows G_3 and G_4. The four main dependent variables are $x_{1,1}$, $x_{2,1}$, $x_{3,2}$ and $x_{4,2}$, and the other $x_{j,K}$ are fixed at zero. With the constraint that the population sums to 1, there are only three degrees of freedom, and we will impose some symmetry on the parameter settings to make a nearly complete analysis possible. Since there are only two UGs, we may restrict our attention to y_1 as $y_2 = 1 - y_1$. This system turns out to exhibit dominance, exclusion, and coexistence, and the remainder of this section is devoted to locating all possible fixed points, using linear stability analysis, and determining which parameter settings show which behavior.

5.3.1. Parameter settings. In this section, we will assume that the payoff matrix B is based on a grammar similarity matrix A, where $A_{i,j}$ is the probability that a sentence spoken at random from G_i can be parsed by a speaker of G_j, and $B = (A + A^T)/2$. A fully general symmetric B matrix for four unambiguous grammars would have six free parameters, as it has 1s down the diagonal and satisfies $B^T = B$ for a total of six degrees of freedom. We therefore make the following simplifying assumption on the form of B:

$$B = \begin{pmatrix} 1 & b_1 & b_2 & b_2 \\ b_1 & 1 & b_2 & b_2 \\ b_2 & b_2 & 1 & b_1 \\ b_2 & b_2 & b_1 & 1 \end{pmatrix} \tag{5.3.1}$$

The parameter b_1 represents how compatible each UG is with itself. A high value indicates that the two possible grammars are very similar and a low value indicates that they are very different. The parameter b_2 represents how compatible the two UGs are with each other. A high value indicates that they are similar, and a low value means they are very different.

A fully general Q matrix would have the form

$$Q_{i,j,1} = \begin{pmatrix} q_1 & 1-q_1 & 0 & 0 \\ 1-q_2 & q_2 & 0 & 0 \\ * & * & * & * \\ * & * & * & * \end{pmatrix}, Q_{i,j,2} = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ 0 & 0 & q_3 & 1-q_3 \\ 0 & 0 & 1-q_4 & q_4 \end{pmatrix}, \tag{5.3.2}$$

with four free parameters. Some entries of Q are filled by $*$ because they are always multiplied by some $x_{j,K}$ that is restricted to be 0, hence, their exact value does not influence the dynamical system. As an additional simplification, we will assume

$$q_2 = q_1 \text{ and } q_4 = q_3,$$

and the remainder of this section will focus on the learning parameters q_1 and q_3, which are the reliabilities of the learning algorithms of U_1 and U_2 respectively.

5.3.2. New coordinates. To make this dynamical system easier to analyze, we will change coordinates so that its symmetry is more readily apparent. The original variables $x_{j,K}$ will be called *simplex coordinates*. These three new variables will be called *box coordinates*:

$$\begin{aligned} r &= \frac{x_{2,1} - x_{1,1}}{x_{2,1} + x_{1,1}}, -1 \le r \le 1, \\ s &= \frac{x_{4,2} - x_{3,2}}{x_{4,2} + x_{3,2}}, -1 \le s \le 1, \\ z &= x_{1,1} + x_{2,1} = y_1, 0 \le z \le 1. \end{aligned} \tag{5.3.3}$$

The balance between G_1 and G_2 is represented by r. Likewise, the balance between G_3 and G_4 is represented by s. Since $z = y_1$, it represents the fraction of the population with U_1. These three pieces of information are enough to identify all possible population states. The reverse change of coordinates is:

$$
\begin{aligned}
x_{1,1} &= \left(\frac{1-r}{2}\right) z, \\
x_{2,1} &= \left(\frac{1+r}{2}\right) z, \\
x_{3,2} &= \left(\frac{1-s}{2}\right) (1-z), \\
x_{4,2} &= \left(\frac{1+s}{2}\right) (1-z).
\end{aligned} \tag{5.3.4}
$$

It is worth noting that the change of coordinates is singular. It expands the simplex in $x_{j,K}$ coordinates into a box in (r, s, z) by blowing up the edges $X_{1,1}X_{2,1}$ and $X_{3,2}X_{4,2}$ into squares, as illustrated in Figure 5.3.1. This singularity does cause some small problems in the fixed point analysis that follows, but they are essentially cosmetic, and will be pointed out as they arise.

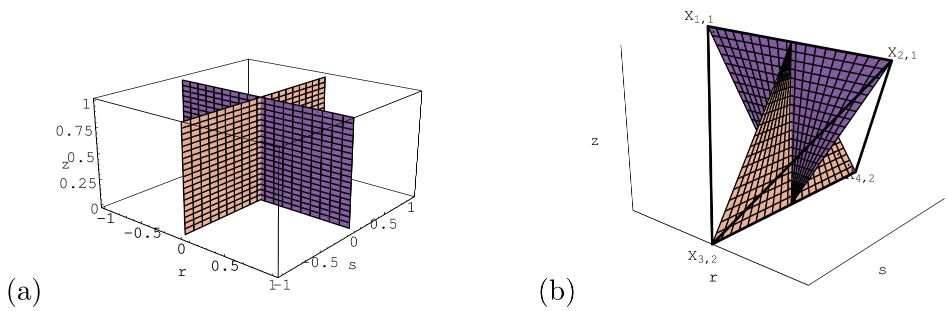


Figure 5.3.1. The singular change of coordinates. (a): The planes given by $r = 0$ and $s = 0$ in (r, s, z) or *box* coordinates. (b): The corresponding sets in simplex coordinates. Note that the top edge of the simplex, where $y_1 = 1$, corresponds to a square in box coordinates. Similarly for the bottom edge, where $y_1 = 0$.

In these new coordinates, the dynamical system (5.2.3) becomes

$$
\begin{aligned}
\dot{r} &= \frac{1}{2} r(-4b_2(1-q_1)(1-z) - (3+b_1-4q_1)z - (1-b_1)r^2 z), \\
\dot{s} &= \frac{1}{2} s(-4b_2(1-q_3)z - (3+b_1-4q_3)(1-z) - (1-b_1)s^2(1-z)), \\
\dot{z} &= \frac{1}{2} z(1-z)((1-b_1)(r^2 z - s^2(1-z)) - (1+b_1-2b_2)(1-2z)).
\end{aligned} \tag{5.3.5}
$$

From this form, it is clear that the planes given by $r = 0$ and $s = 0$ are invariant under the dynamics, as are the sets where $z = 0$ or $z = 1$. Once the fixed points are located, the behavior of the dynamical system will be determined by focusing on these invariant planes.

5.3.3. Locating the fixed points. This system has a lot of fixed points, and to keep them straight, each one will be given a subscript. So for example, by looking at (5.3.5), several fixed points are obvious. If we assume $r = 0$ and $s = 0$, then $\dot{r} = 0$, $\dot{s} = 0$, and $\dot{z}$ is a cubic polynomial in z, so there are exactly these three fixed points:

$$
\begin{aligned}
(r_{T0}, s_{T0}, z_{T0}) &= (0, 0, 1),\\
(r_{B0}, s_{B0}, z_{B0}) &= (0, 0, 0),\\
(r_{C0}, s_{C0}, z_{C0}) &= \left(0, 0, \frac{1}{2}\right).
\end{aligned} \tag{5.3.6}
$$

For reference, the T subscript means "top," B means "bottom," and C means "center." These three fixed points always exist.

The remaining fixed points exist only for certain values of the four parameters. The complete conditions on b_1, b_2, and q_3 for the existence of all fixed points will be derived in Section 5.3.4. The first step is to develop expressions for all the other fixed points, and it turns out that they all involve products of certain repeated linear expressions in q_1 and q_3. For simplicity of notation, we will name these linear expressions[1] as follows:

$$
\begin{aligned}
A &= 1 + b_1 - 2b_2, & &\\
B &= 4b_2 q_3 - 1 - b_1 - 2b_2, & q_B &= \frac{1 + b_1 + 2b_2}{4b_2},\\
C &= 4q_3(1 + b_2) - 3 - b_1 - 4b_2, & q_C &= \frac{3 + b_1 + 4b_2}{4(1 + b_2)},\\
D &= 2q_3 - b_2 - 1, & q_D &= \frac{1 + b_2}{2},\\
E &= 4q_3(1 - b_2) - (1 - b_1), & q_E &= \frac{1 - b_1}{4(1 - b_2)},\\
F &= (2 + 4b_2)q_3 - 2 - b_1 - 3b_2, & q_F &= \frac{2 + b_1 + 3b_2}{2 + 4b_2},\\
G &= 4q_3 - 3 - b_1, & q_G &= \frac{3 + b_1}{4},\\
B' &= 4b_2 q_1 - 1 - b_1 - 2b_2, & q_{B'} &= q_B,\\
C' &= 4q_1(1 + b_2) - 3 - b_1 - 4b_2, & q_{C'} &= q_C,\\
D' &= 2q_1 - b_2 - 1, & q_{D'} &= q_D,\\
E' &= 4q_1(1 - b_2) - (1 - b_1), & q_{E'} &= q_E,\\
G' &= 4q_1 - 3 - b_1, & q_{G'} &= q_G.
\end{aligned} \tag{5.3.7}
$$

The signs of these expressions will be important, as most of the fixed points have a coordinate that is a square-root of some product of them, so it will only exist when those products are positive. These expressions were selected to follow the convention that they are negative if q_3 or q_1 is small. The value of q_3 for which $B = 0$ will be called a *turning point* and is

[1] It simplifies the notation greatly to use single letters for these expressions, however, we are starting to run out of letters. Rather than choose an odd collection of the remainder of the alphabet, we will use A through G for these constants and no longer refer in this section to the matrices A and B or the vector F as used in as used in Section 5.2 as they are no longer needed, and the constant G here can be distinguished from grammar G_j by the presence of a subscript.

denoted q_B. Turning points for the other expressions (except A) are defined similarly, and they will be used extensively in determining which fixed points exist for which parameter values.

If we look for additional fixed points with $z = 0$, then $\dot{z} = 0$ automatically, and $\dot{r} = -2b_2(1 - q_1)$ which forces $r = 0$. We get a cubic polynomial

$$\dot{s} = \frac{1}{2}s(-(3 + b_1 - 4q_3) - (1 - b_1)s^2),$$

one root of which is $s = 0$ which has already been covered. The other two roots yield two more fixed points,

$$\begin{aligned} (r_{B1}, s_{B1}, z_{B1}) &= \left(0, \sqrt{\frac{G}{1 - b_1}}, 0\right), \\ (r_{B2}, s_{B2}, z_{B2}) &= \left(0, -\sqrt{\frac{G}{1 - b_1}}, 0\right), \end{aligned} \tag{5.3.8}$$

that exist when $G \geq 0$.

Similarly, the assumption $z = 1$ yields two more fixed points that exist when $G' \geq 0$:

$$\begin{aligned} (r_{T1}, s_{T1}, z_{T1}) &= \left(\sqrt{\frac{G'}{1 - b_1}}, 0, 1\right), \\ (r_{T2}, s_{T2}, z_{T2}) &= \left(-\sqrt{\frac{G'}{1 - b_1}}, 0, 1\right). \end{aligned} \tag{5.3.9}$$

Assuming $r = 0, s \neq 0, z \neq 0, z \neq 1$, there are two more fixed points which will be indicated by subscripts $S1$ and $S2$. To find them, we cancel some factors from the equations $\dot{s} = 0$ and $\dot{z} = 0$ to arrive at the system

$$\begin{aligned} 0 &= (1 - b_1)s^2(1 - z) + 4b_2(1 - q_3)z + (3 + b_1 - 4q_3)(1 - z), \\ 0 &= -(1 - b_1)s^2(1 - z) - (1 + b_1 - 2b_2)(1 - 2z). \end{aligned}$$

The sum of these two equations is linear in z with no s, and has a unique root. Substituting this value of z back into either of the preceding equations yields two solutions for s, so we find a total of two new fixed points:

$$\begin{aligned} (r_{S1}, s_{S1}, z_{S1}) &= \left(0, \sqrt{\frac{AC}{(1 - b_1)(-B)}}, \frac{2D}{E}\right), \\ (r_{S2}, s_{S2}, z_{S2}) &= \left(0, -\sqrt{\frac{AC}{(1 - b_1)(-B)}}, \frac{2D}{E}\right). \end{aligned} \tag{5.3.10}$$

It will be necessary to know the parameter values for which each fixed point exists, so it is useful to observe that

$$1 - z_{S1} = 1 - \frac{2D}{E} = -\frac{B}{E}.$$

The $S1$ and $S2$ fixed points only exist when $z_{S1} > 0$ and $1 - z_{S1} > 0$.

Similarly, by assuming $s = 0$, we find two more fixed points with associated subscripts $R1$ and $R2$:

$$\begin{aligned}(r_{R1}, s_{R1}, z_{R1}) &= \left(\sqrt{\frac{AC'}{(1-b_1)(-B')}}, 0, 1 - \frac{2D'}{E'}\right),\\ (r_{S2}, s_{S2}, z_{S2}) &= \left(-\sqrt{\frac{AC'}{(1-b_1)(-B')}}, 0, 1 - \frac{2D'}{E'}\right).\end{aligned} \tag{5.3.11}$$

There are four more fixed points that come from assuming that r, s, and z are nonzero, and $z \neq 1$. These will be given subscripts $C1$, $C2$, $C3$, and $C4$. Setting the dynamical system in (5.3.5) to zero and canceling factors of r, s, z and $1 - z$, we get a system of equations

$$0 = -4b_2(1-q_1)(1-z) - (3 + b_1 - 4q_1)z - (1-b_1)r^2 z, \tag{5.3.12a}$$

$$0 = -4b_2(1-q_3)z - (3 + b_1 - 4q_3)(1-z) - (1-b_1)s^2(1-z), \tag{5.3.12b}$$

$$0 = (1-b_1)s^2(1-z) - (1-b_1)r^2 z + (1 + b_1 - 2b_2)(1-2z). \tag{5.3.12c}$$

Taking (5.3.12b) plus (5.3.12c) minus (5.3.12a) yields a linear equation in z with no r or s which yields the unique root

$$z_{C1} = z_{C2} = z_{C3} = z_{C4} = \frac{1 - 2q_3 - b_2(1-2q_1)}{2(1-b_2)(1-q_1-q_3)}. \tag{5.3.13}$$

Substituting this value back into the system allows us to solve for the remaining coordinates,

$$\begin{aligned} r_{C1} &= \sqrt{\frac{(4q_1 - 3 - b_1)z_{C1} - 4b_2(1-q_1)(1-z_{C1})}{(1-b_1)z_{C1}}},\\ r_{C4} &= r_{C1}, r_{C2} = r_{C3} = -r_{C1},\\ s_{C1} &= \sqrt{\frac{(4q_3 - 3 - b_1)(1-z_{C1}) - 4b_2(1-q_3)z_{C1}}{(1-b_1)(1-z)}},\\ s_{C2} &= s_{C1}, s_{C3} = s_{C4} = -s_{C1}.\end{aligned} \tag{5.3.14}$$

For reference, the labels of the fixed points with $r = 0$ are illustrated in Figure 5.3.2, which is the cross-section of the simplex corresponding to $r = 0$.

5.3.4. Linear stability analysis in the invariant planes. Now that we have a list of all possible fixed points, what remains is to find the parameter settings for which they exist and whether they are sinks, sources, or saddles. The fact that the dynamical system has two invariant planes, given by $r = 0$ and $s = 0$ respectively, simplifies the analysis as we can focus on one plane at a time, then assemble the pieces into complete phase portraits. The phase portraits for the $r = 0$ plane will be developed in three groups, based on the ordering of b_2, b_2^2, and $(1 + b_1)/2$. The bifurcation values of q_3 are determined by looking at the turning points of the linear expressions A through G from (5.3.7). The order in which they occur as q_3 increases is determined by several propositions relating them to the ordering of b_2, b_2^2 and $(1 + b_1)/2$. The type of bifurcation that occurs at each of these values

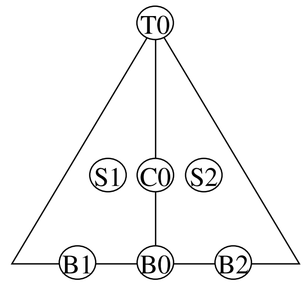


Figure 5.3.2. Labeled fixed points in the $r = 0$ plane. These are the approximate positions of these fixed points when they exist. The triangle is a cross-section of the simplex as shown in Figure 5.3.1.

is determined by which fixed points are non-hyperbolic there. The analysis for the $s = 0$ plane is similar and will not be presented here.

5.3.4.1. *Local analysis of the fixed points.* Since we are working in box coordinates with r fixed at zero, each fixed point has a 2 by 2 Jacobian matrix, which will be expressed in terms of the unit vectors $\{\hat{s}, \hat{z}\}$ which point rightward and upward, respectively.

At $B0$, the Jacobian matrix is diagonal, and the eigenvalues can be read off as $G/2$ for the eigenvector $\hat{s}$ and $-A/2$ for the eigenvector $\hat{z}$. For $B1$ and $B2$, the Jacobian is triangular, and the eigenvalues are $-G$ and $-D$. For $q_3 < q_G$, only $B0$ exists on the bottom, and it is stable in the $\hat{s}$ direction. As q_3 increases through q_G, there is a pitchfork bifurcation resulting in the creation of $B1$ and $B2$. For the central fixed point $C0$, the Jacobian is also diagonal, with eigenvalues $C/4$ for $\hat{s}$, and $A/4$ for $\hat{z}$. The top fixed point $T0$ also has a diagonal Jacobian, with eigenvalues $G/2$ for $\hat{s}$ and $-A/2$ for $\hat{z}$. The $\hat{s}$ direction does not affect the stability of $T0$ because the singular change of coordinates from (r, s, z) to simplex variables collapses the $\hat{s}$ direction.

The $S1$ and $S2$ fixed points do not exist for all values of the parameters. When they do exist, they do not have triangular Jacobians. Rather than determine their eigenvalues directly, we use the fact that the stability of a fixed point in a planar dynamical system can be determined from the trace and determinant of its Jacobian:

Lemma 5.3.1. *Let d and t be the determinant and trace of the Jacobian matrix at a fixed point $\bar{x}$ in a planar dynamical system. If $d < 0$, then $\bar{x}$ is a saddle. If $d > 0$ and $t < 0$, then $\bar{x}$ is a sink. If $d > 0$ and $t > 0$, then $\bar{x}$ is a source.*

Proof. See [70] p. 137. □

The fixed points $S1$ and $S2$ happen to have the same determinant, $ABCD/E^2$, and the same trace, $-AF/E$. It is worth noting that $S1$ and $S2$ are on the bottom line when $z_{S1} = z_{S2} = 0$, which happens when $D = 0$ and implies that they coincide with $B1$ and $B2$. This means that as q_3 increases through q_D, D must change sign, and $S1$ and $S2$ must pass through $B1$ and $B2$ in a pair of transcritical bifurcations. Likewise, $S1$ and $S2$ are on the vertical line $s = 0$ when $s_{S1} = s_{S2} = 0$. This happens when $C = 0$ or $A = 0$. The case when $A = 0$ is degenerate and non-generic, so it will not be analyzed here. If $C = 0$, then $S1$ and $S2$ coincide with $C0$, which implies a pitchfork bifurcation.

5.3.4.2. *Propositions for turning points.* To determine which fixed points exist for which parameter values and what their stabilities are, we will use the following propositions to determine the signs of A, B, C, D, E, F, and G. Each of these expression except A can be thought of as a linear function of q_3 with a single turning point at which the expression changes sign. These propositions determine the order of those turning points from the ordering of b_2, b_2^2, and $(1 + b_1)/2$.

Proposition 5.3.2. $\operatorname{sgn} q_B - 1 = \operatorname{sgn} A$.

Proof.

$$\begin{aligned}
\operatorname{sgn} q_B - 1 &= \frac{1 + b_1 + 2b_2}{4b_2} - 1 \\
&= \operatorname{sgn} 1 + b_1 + 2b_2 - 4b_2 \\
&= \operatorname{sgn} 1 + b_1 - 2b_2.
\end{aligned}$$

□

Proposition 5.3.3. $\operatorname{sgn} q_B - q_F = \operatorname{sgn} 1 + b_1 - b_2^2$.

Proof.

$$\begin{aligned}
\operatorname{sgn} q_B - q_F &= \operatorname{sgn} \frac{1 + b_1 + 2b_2}{4b_2} - \frac{2 + b_1 + 3b_2}{2 + 4b_2} \\
&= \operatorname{sgn}(1 + b_1 + 2b_2)(1 + 2b_2) - 2b_2(2 + b_1 + 3b_2) \\
&= \operatorname{sgn} 1 + b_1 - b_2^2.
\end{aligned}$$

□

Proposition 5.3.4. $\operatorname{sgn} q_F - q_C = \operatorname{sgn} 1 + b_1 - 2b_2^2$.

Proof.

$$\begin{aligned}
\operatorname{sgn} q_F - q_C &= \operatorname{sgn} \frac{2 + b_1 + 3b_2}{2 + 4b_2} - \frac{3 + b_1 + 4b_2}{4(1 + b_2)} \\
&= \operatorname{sgn} 2(1 + b_2)(2 + b_1 + 3b_2) - (3 + b_1 + 4b_2)(1 + 2b_2) \\
&= \operatorname{sgn} 1 + b_1 - 2b_2^2.
\end{aligned}$$

□

Proposition 5.3.5. $\operatorname{sgn} q_C - q_D = \operatorname{sgn} 1 + b_1 - 2b_2^2$.

Proof.

$$\begin{aligned}
\operatorname{sgn} q_C - q_D &= \operatorname{sgn} \frac{3 + b_1 + 4b_2}{4(1 + b_2)} - \frac{1 + b_2}{2} \\
&= \operatorname{sgn} 3 + b_1 + 4b_2 - 2(1 + b_2)^2 \\
&= \operatorname{sgn} 1 + b_1 - 2b_2^2.
\end{aligned}$$

□

Proposition 5.3.6. $\operatorname{sgn} q_D - q_E = \operatorname{sgn} 1 + b_1 - 2b_2^2$.

Proof.

$$\begin{aligned}\operatorname{sgn} q_D - q_E &= \frac{1+b_2}{2} - \frac{1-b_1}{4(1-b_2)} \\ &= \operatorname{sgn} 2(1-b_2^2) - (1-b_1) \\ &= \operatorname{sgn} 1 + b_1 - 2b_2^2.\end{aligned}$$

□

Proposition 5.3.7. $\operatorname{sgn} q_G - q_D = \operatorname{sgn} A$.

Proof.

$$\begin{aligned}\operatorname{sgn} q_G - q_D &= \operatorname{sgn} \frac{3+b_1}{4} - \frac{1+b_2}{2} \\ &= \operatorname{sgn} 3 + b_1 - 2 - 2b_2 \\ &= \operatorname{sgn} 1 + b_1 - 2b_2.\end{aligned}$$

□

Proposition 5.3.8. $q_C \geq q_G$.

Proof.

$$\begin{aligned}\operatorname{sgn} q_C - q_G &= \operatorname{sgn} \frac{3+b_1+4b_2}{4(1+b_2)} - \frac{3+b_1}{4} \\ &= \operatorname{sgn} 3 + b_1 + 4b_2 - (3+b_1)(1+b_2) \\ &= \operatorname{sgn} b_2(1-b_1) = 0 \text{ or } 1.\end{aligned}$$

Note that the inequality is strict if we assume strict inequalities on b_1 and b_2. That is, if $0 < b_1 < 1$ and $0 < b_2 < 1$, then $q_C > q_G$. □

We would like to identify the bifurcations that happen as q_3 increases from 0 to 1. From these propositions, it is clear that the order in which the various expressions A through G change sign depends on the relative sizes of b_1 and b_2. In particular, we have three cases depending on where $(1+b)/2$ lies with respect to b_2 and b_2^2. We will break up the analysis into these three cases and analyze them separately.

5.3.4.3. *First group of phase portraits.* For now, assume that

$$b_2 > \frac{1+b_1}{2} > b_2^2, \tag{5.3.15}$$

which is a case of moderate similarity between the two universal grammars. In what follows, we will discuss what happens to all the fixed points in the $r = 0$ plane as q_3 increases from 0 to 1. With these assumptions, $A < 0$, so the central fixed point $C0$ is always vertically stable. Also, we have the ordering of turning points $q_E < q_D < q_C < q_B < 1$ and $q_G < q_D$.

This information is enough to conclude that the $S1$ and $S2$ fixed points exist when $q_D < q_3 < q_C$, meaning that s_{S1} and s_{S2} are real valued, and $0 < z_{S1} < 1$. That range for q_3 is determined by looking at the table in Figure 5.3.3 for regions where s_{S1}^2, z_{S1} and

$1 - z_{S1}$ are all positive. The signs of these expressions are determined from the signs of A through G which depend on whether q_3 is above or below the corresponding turning point. When the $S1$ and $S2$ fixed points exist, the determinant of the Jacobian there is negative, so they are saddles.

q_3	$s_{S1}^2 = -\dfrac{AC}{B}$	$z_{S1} = \dfrac{2D}{E}$	$1 - z_{S1} = -\dfrac{B}{E}$	$\det = \dfrac{ABCD}{E^2}$
1				
	$+$	$+$	$-$	
q_B				
	$-$	$+$	$+$	
q_C				
	$+$	$+$	$+$	$-$
q_D				
	$+$	$-$	$+$	
q_E				
	$+$	$+$	$-$	
0				

Figure 5.3.3. Sign table for the first set of phase portraits, assuming $b_2 > (1+b_1)/2 > b_2^2$. The value of q_3 increases from 0 at the bottom to 1 at the top.

The $B1$ and $B2$ fixed points exist for $q_3 > q_G$, and since $q_G < q_D$, they always exist when $S1$ and $S2$ exist. Furthermore, they are saddles while $q_G < q_3 < q_D$, but they become sinks as q_3 passes through q_D. In fact, $z_{S1} = z_{S2} = 0$ at $q_3 = q_D$, which means that $S1$ and $S2$ pass through $B1$ and $B2$ in transcritical bifurcations.

At $q_3 = q_C$, the fixed points $S1$ and $S2$ collide with $C0$ because $s_{S1}^2 = s_{S2}^2 = 0$. Since they wink out of existence at q_C, this is a pitchfork bifurcation.

The possible phase portraits for the assumptions in (5.3.15) are depicted in Figure 5.3.4. The pictures here are produced by working in the plane $r = 0$ with (s, z) coordinates that cover a rectangular cross section in the box-shaped phase space of Figure 5.3.1 (a), then transforming the result into simplex coordinates yielding a triangular cross section of the simplex phase space of Figure 5.3.1 (b). The stability of each fixed point is drawn with respect to the two eigenvalues of its Jacobian whose eigenvectors lie in that plane. In the full simplex, each of these fixed points has a third eigenvalue whose eigenvector points out of the restricted phase space. In Section 5.3.6 phase portraits are drawn for the full phase space and some fixed points depicted here are drawn there with a different stability symbol because that third eigenvalue becomes relevant.

An additional complexity comes from the fact that the change of coordinates is singular. In box coordinates, the fixed points on the top line of the simplex are often stable in the $\hat{s}$ direction, but the entire plane $z = 1$ collapses into a line in simplex coordinates and the $\hat{s}$ direction disappears. This collapse appears in the plane $r = 0$ in that the line $z = 1$ collapses down to a point (the apex of the triangle in these figures) and the $\hat{s}$ direction disappears there. So, in the following diagrams, the eigenvalue associated with the eigenvector $\hat{s}$ is ignored when determining the stability of $T0$. It sometimes shows up as a saddle in (s, z) coordinates but is drawn as a source in the simplex because no orbit in the simplex converges to it.

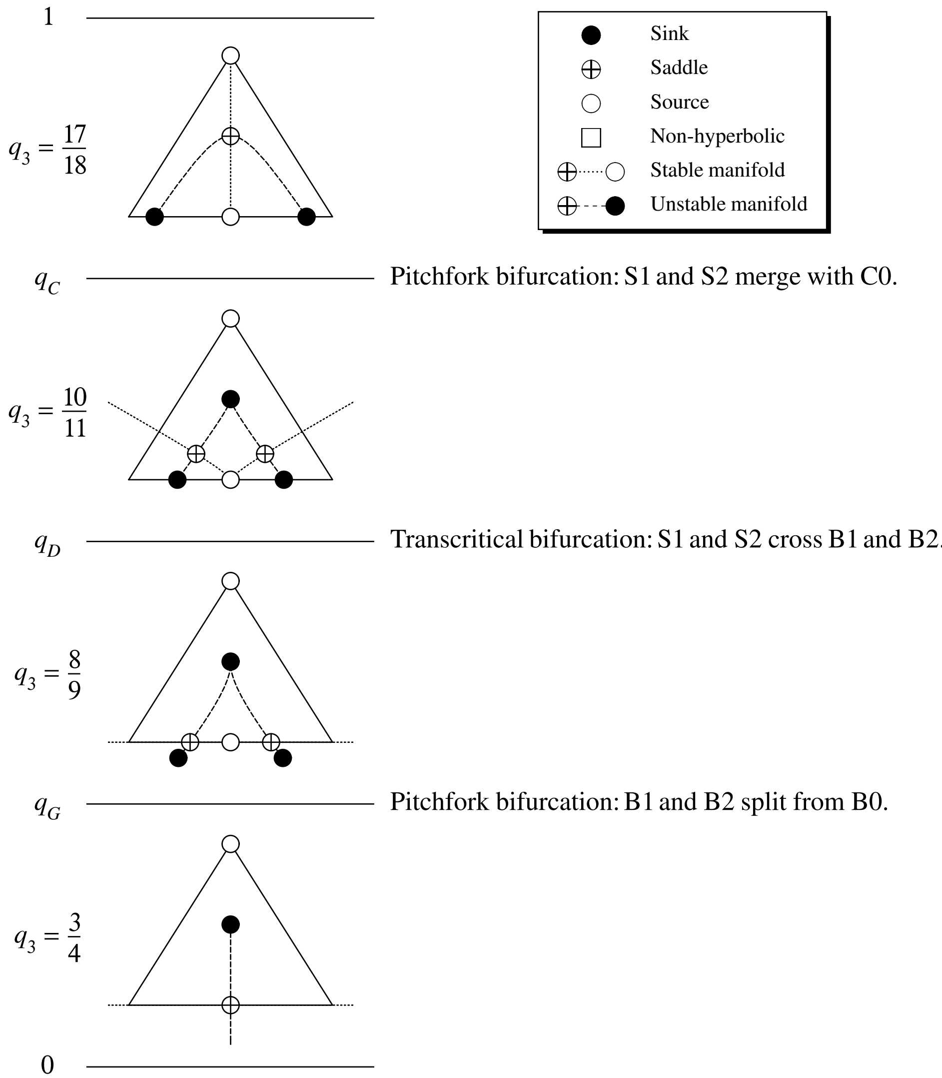


Figure 5.3.4. Fist set of possible phase portraits for the plane $r = 0$ assuming $b_2 > (1 + b_1)/2 > b_2^2$. The value of q_3 increases from 0 at the bottom to 1 at the top. There is one phase portrait for each range of q_3 between important turning points. See text for a note about the $T0$ fixed points, which are saddles in (s, z) coordinates and sources in simplex coordinates. For these pictures, $b_1 = 1/2$ and $b_2 = 4/5$.

5.3.4.4. *Second group of phase portraits.* For this group of phase portraits, we assume

$$b_2 > b_2^2 > \frac{1+b_1}{2}. \tag{5.3.16}$$

Here, the two UGs are highly similar. With these assumptions, we again have $A < 0$, so the central fixed point $C0$ is always vertically stable. This time, $1 + b_1 - 2b_2^2 < 0$, so the ordering of turning points is $q_B < q_F < q_C < q_D < q_E$ and $q_G < q_C$. The sign table is given in Figure 5.3.5, and indicates that the $S1$ and $S2$ fixed points exist exactly when $q_C < q_3 < q_D$ and they are sinks. At $q_3 = q_C$, both $S1$ and $S2$ coincide with $C0$, so there is a pitchfork bifurcation there. Since $q_G < q_C$, the fixed points $B1$ and $B2$ already exist when $S1$ and $S2$ come into existence. At $q_3 = q_D$, $S1$ and $S2$ pass through $B1$ and $B2$ because $z_{S1} = z_{S2} = 0$, so there is a pair of transcritical bifurcations there.

q_3	$s_{S1}^2 = -\frac{AC}{B}$	$z_{S1} = \frac{2D}{E}$	$1 - z_{S1} = -\frac{B}{E}$	$\det = \frac{ABCD}{E^2}$	$\mathrm{tr} = -\frac{AF}{E}$
1					
	+	+	−		
q_E					
	+	−	+		
q_D					
	+	+	+	+	−
q_C					
	−	+	+		
q_F					
	−	+	+		
q_B					
	+	+	−		
0					

Figure 5.3.5. Sign table for the second set of phase portraits, assuming $b_2 > b_2^2 > (1+b_1)/2$. The value of q_3 increases from 0 at the bottom to 1 at the top.

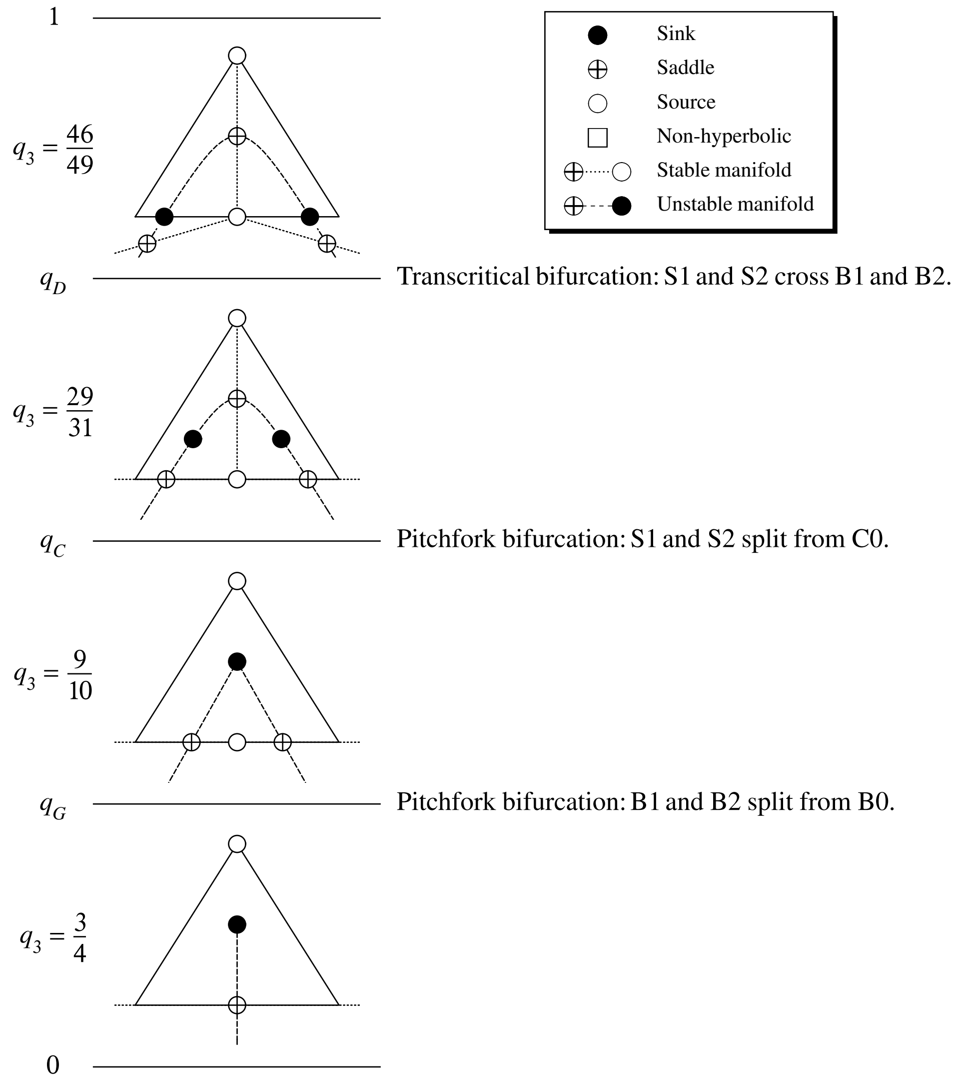


Figure 5.3.6. Second set of possible phase portraits for the plane $r = 0$ assuming $b_2 > b_2^2 > (1 + b_1)/2$. The value of q_3 increases from 0 at the bottom to 1 at the top. There is one phase portrait for each range of q_3 between important turning points. See text for a note about the $T0$ fixed points, which are saddles in (s, z) coordinates and sources in simplex coordinates. For these pictures, $b_1 = 1/2$ and $b_2 = 7/8$.

5.3.4.5. *Third group of phase portraits.* For this final group of phase portraits, we assume

$$\frac{1+b_1}{2} > b_2 > b_2^2, \tag{5.3.17}$$

so the two UGs are dissimilar. Here, $A > 0$, so $C0$ is vertically unstable and the order of turning points is $q_E < q_D < q_G < q_C < q_F < 1 < q_B$. The sign table for $S1$ and $S2$ is given in Figure 5.3.7, indicates that these two fixed points only exist when $q_3 > q_C$, and they are always saddles. In this case, there is no transcritical bifurcation where $S1$ and $S2$ cross $B1$ and $B2$. The phase portraits are displayed in Figure 5.3.8.

q_3	$s_{S1}^2 = -\dfrac{AC}{B}$	$z_{S1} = \dfrac{2D}{E}$	$1 - z_{S1} = -\dfrac{B}{E}$	$\det = \dfrac{ABCD}{E^2}$
1				
	+	+	+	−
q_C				
	−	+	+	
q_D				
	−	−	+	
q_E				
	−	+	−	
0				

Figure 5.3.7. Sign table for the third set of phase portraits, assuming $(1+b_1)/2 > b_2 > b_2^2$. The value of q_3 increases from 0 at the bottom to 1 at the top.

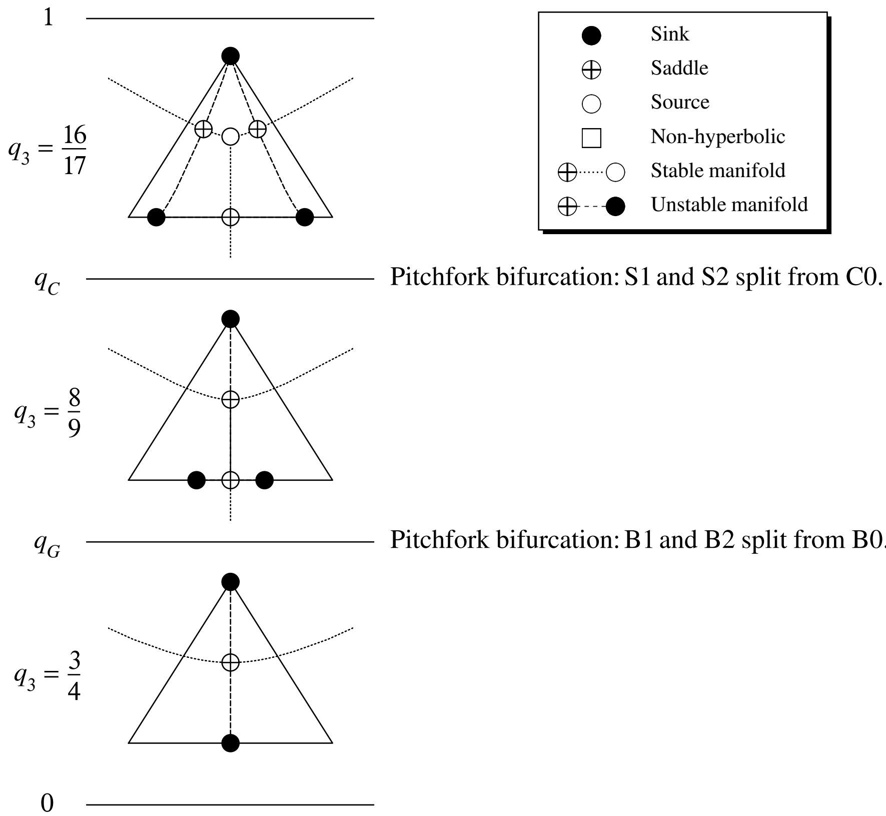


Figure 5.3.8. Third set of possible phase portraits for the plane $r = 0$ assuming $(1 + b_1)/2 > b_2 > b_2^2$. The value of q_3 increases from 0 at the bottom to 1 at the top. There is one phase portrait for each range of q_3 between important turning points. For these pictures, $b_1 = 1/2$ and $b_2 = 1/3$.

5.3.5. Out-of-plane stability and the remaining four fixed points. The goal of this analysis is to determine parameter values for which the overall dynamical system exhibits dominance, exclusion, and coexistence. Now that we have a complete picture of what happens in the invariant planes $r = 0$ and $s = 0$ for different parameter values, the next step is to determine the out-of-plane stabilities of the fixed points in these invariant planes. The analysis for the plane $r = 0$ will be carried out here. The $s = 0$ plane is similar, and will not be explicitly analyzed.

The dynamical system has the property that the Jacobian matrix for any point where $r = 0$ has the form

$$\begin{pmatrix} \frac{\partial \dot{r}}{\partial r}\Big|_{r=0} & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{pmatrix}$$

where the $*$ entries may be non-zero. Any matrix of this form has an eigenvector $(1, 0, 0)$ purely in the $\hat{r}$ direction whose associated eigenvalue is

$$\begin{aligned} \lambda_r &= \frac{\partial \dot{r}}{\partial r}\bigg|_{r=0} \\ &= -2b_2(1-q_1) + \left(\frac{4q_1 - 3 - b_1}{2} + 2b_2(1-q_1)\right) z. \end{aligned} \tag{5.3.18}$$

This eigenvalue determines the out-of-plane stability of a fixed point in the plane $r = 0$. It has a single sign change at $z = z_P$ where

$$z_P = \frac{4b_2(1-q_1)}{4q_1(1-b_2) + 4b_2 - 3 - b_1}. \tag{5.3.19}$$

If $z = 0$, then $\lambda_r = -2b_2(1-q_1) < 0$. What this means is that there is a horizontal line on the plane $r = 0$ at $z = z_P$, and any fixed points below this line are stable in the $\hat{r}$ direction and any fixed points above this line are unstable in the $\hat{r}$ direction. If the other parameters are held fixed and q_1 increases from 0 to 1, then z_P sweeps through the plane from top to bottom. The line passes through the $C0$ fixed point when $z_P = 1/2$, which occurs at $q_1 = q_C$, corresponding to a pitchfork bifurcation in the plane $s = 0$. When the line passes through $S1$ and $S2$, the fixed points $C1$ and $C2$ coincide with $S1$, and $C3$ and $C4$ coincide with $S2$. To see this, observe that if $z_P = z_{S1}$, then

$$q_1 = \frac{-3 - b_2 + b_2^2(4 - 8q_3) + 6q_3 + b_1(-1 + b_2 + 2q_3)}{2(-2 + (-1 + b_1)b_2 + b_2^2(2 - 4q_3) + 4q_3)}.$$

With some further simplification and substitution this value for q_1, it follows that $z_{C1} = z_{S1}$ and $r_{C1} = 0$, so $C1$ coincides with $S1$. The collisions of the other fixed points follow symmetrically. Thus, there is a symmetric pair of pitchfork bifurcations when $z_P = z_{S1}$.

The expressions for the four fixed points $C1$, $C2$, $C3$, and $C4$ are fairly complicated and are not conducive to further symbolic results. So in the next section, we will pick a number of specific values for the four parameters and assemble the phase portrait from the results so far.

5.3.6. Some sample phase portraits. For fixed b_1, b_2, q_1, and q_3, the locations and stabilities of all fixed points can be determined numerically, and the results drawn in a 3-dimensional phase portrait. As stated earlier, the calculations are easiest in box coordinates, but the fixed points on the bottom and top edges of the simplex have a misleading eigenvalue corresponding to the eigenvector ($\hat{r}$ or $\hat{s}$, respectively) that is introduced by the singular transformation from the simplex to box coordinates. So, in the following pictures, those eigenvalues are ignored when selecting a stability symbol for the fixed points on the top and bottom edges.

First, Figure 5.3.9 shows some complete phase portraits for the first set of parameter values where the two UGs are moderately similar,

$$b_2 > \frac{1+b_1}{2} > b_2^2.$$

Here, q_3 is held constant and q_1 increases, illustrating some bifurcations in the invariant plane $s = 0$. The bottom picture illustrates stable coexistence, as there is a sink in the middle of the simplex where both universal grammars take up half the population. The middle picture illustrates a case where in addition to stable coexistence, U_1 can take over the population in a stable manner. The two saddles on the top line in the bottom picture become sinks in the middle picture through a pair of transcritical bifurcations, enabling the model to exhibit this new behavior. The top picture, which occurs after the two interior saddles, $R1$ and $R2$, collide with the sink in the middle, $C0$. This bifurcation destroys the coexistence equilibrium, leaving a situation where U_1 dominates.

The second group of phase portraits, with parameter values

$$b_2 > b_2^2 > \frac{1+b_1}{2},$$

exhibits only an asymmetric form of stable coexistence. The pictures in Figure 5.3.10 illustrate several different cases. Instead of a sink at $C0$, the model has sinks at $R1$, $R2$, $S1$, and $S2$, so both universal grammars coexist stably, but asymmetrically. Here, the two UGs are highly similar.

For the third group of parameter values,

$$\frac{1+b_1}{2} > b_2 > b_2^2,$$

the two UGs are dissimilar, and all of the interior fixed points on the invariant planes are saddles or sources. The remaining four fixed points $C1$, $C2$, $C3$, and $C4$ appear to be saddles in general. They are definitely saddles for the cases depicted in Figure 5.3.11. All of these pictures are different kinds of exclusion, in which the population is eventually taken over by either U_1 or U_2. Each sink has a basin of attraction, and the boundaries of these basins are probably formed from the stable manifolds of the interior saddle points. The different configurations of saddle points would then correspond to alternative ways of forming boundaries between the basins.

5.3.7. Discussion. The case of the language dynamical equation examined in this section exhibits dominance, competitive exclusion, and coexistence of universal grammars. Furthermore, all three kinds of behavior appear to be generic. The complete phase portraits shown

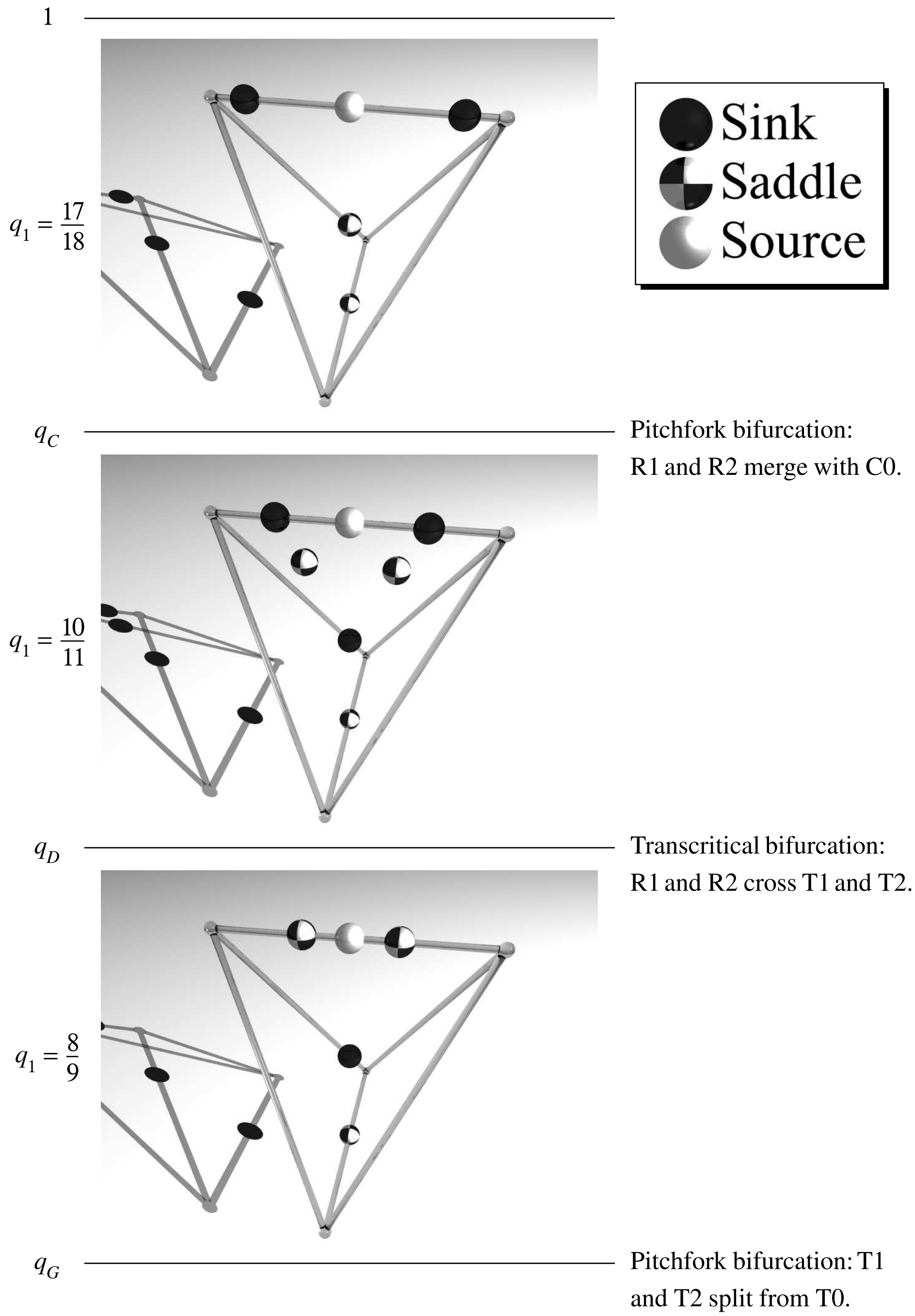


Figure 5.3.9. Complete 3-dimensional phase portraits. For these pictures, $b_1 = 1/2$, $b_2 = 4/5$, and $q_3 = 3/4$. The value of q_1 increases from bottom to top.

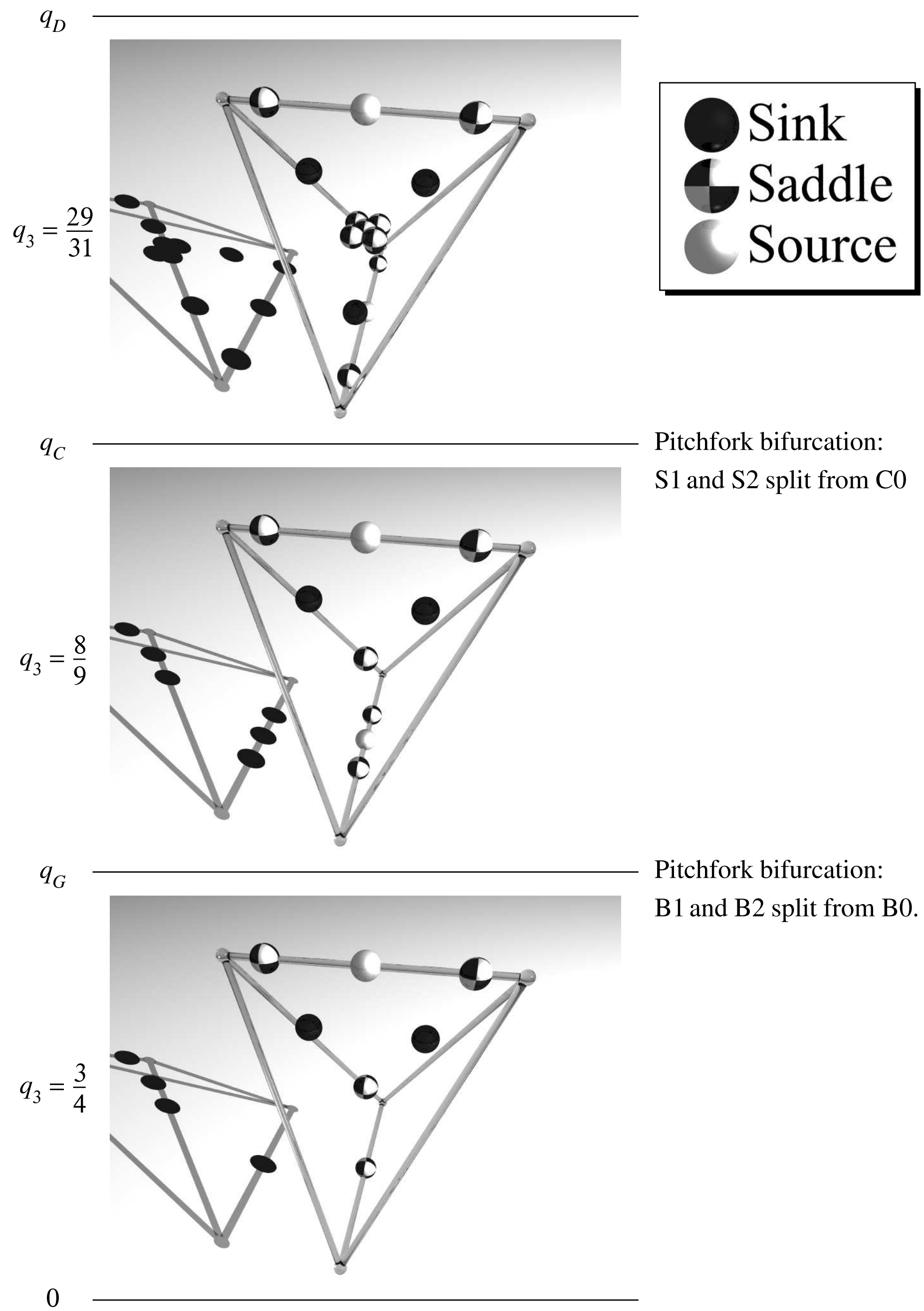


Figure 5.3.10. Complete 3-dimensional phase portraits. For these pictures, $b_1 = 1/2$, $b_2 = 7/8$, and $q_1 = 29/31$. The value of q_3 increases from bottom to top.

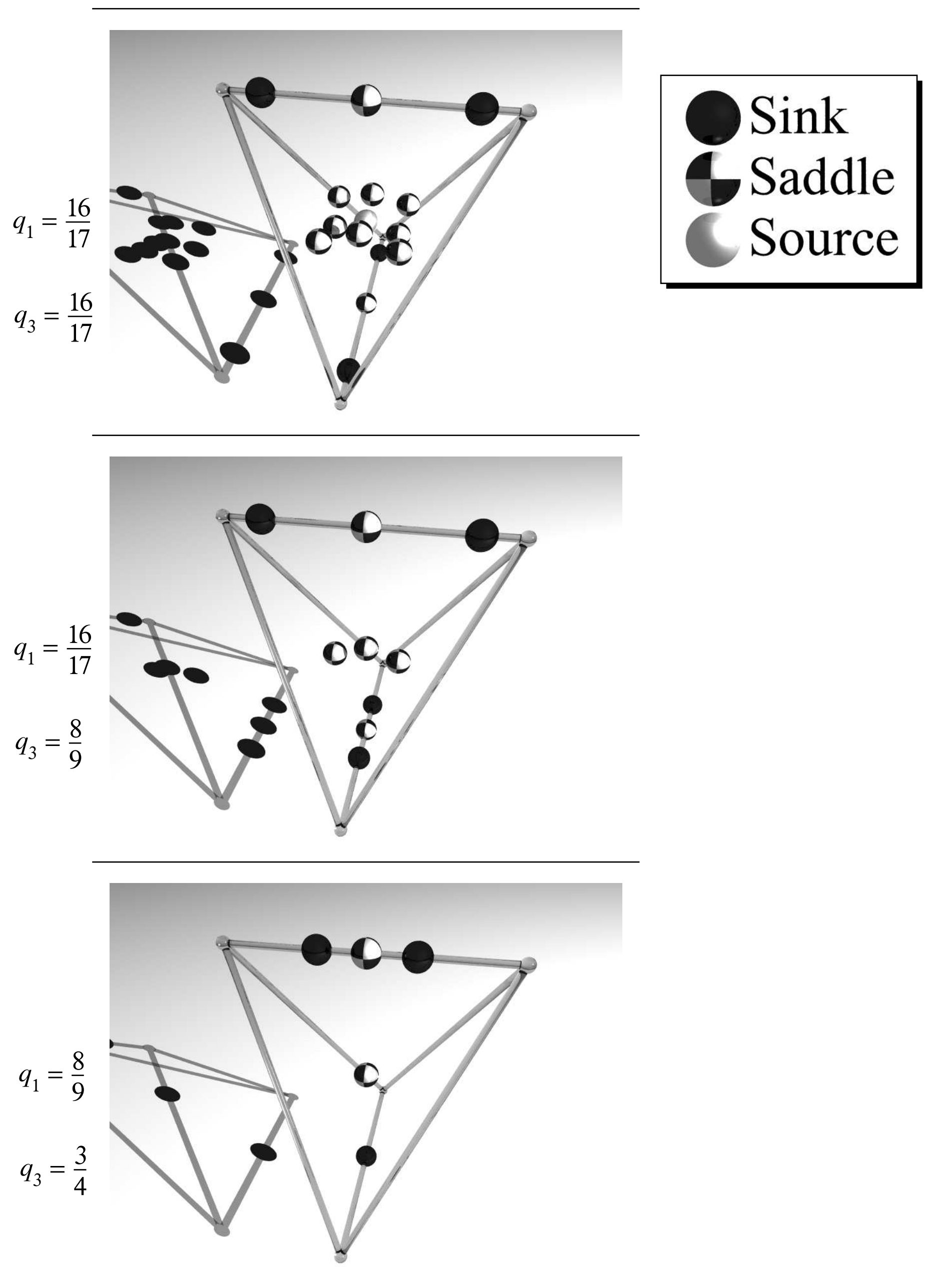
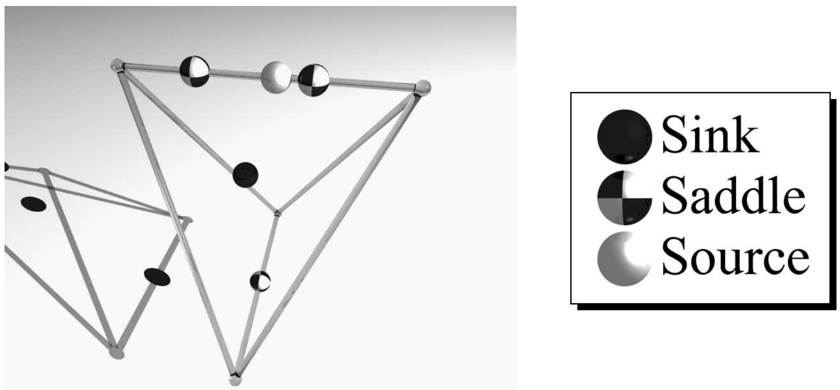


Figure 5.3.11. Complete 3-dimensional phase portraits. For these pictures, $b_1 = 1/2$ and $b_2 = 1/3$, and q_1 and q_3 both vary.

so far appear to be structurally stable, meaning that all fixed points continue to exist with the same stability type under perturbations of the parameters. In particular, the existence of stable fixed points where the two universal grammars coexist is not an artifact of the symmetric parameter settings. To illustrate this fact, Figure 5.3.12 shows a perturbation of the center phase portrait in Figure 5.3.9. The stable fixed point inside the simplex persists despite the broken symmetry. UG is normally thought of as a universal trait, shared by all human beings, but according to this model, it is possible for multiple UGs to exist stably within a population.



$$B = \begin{pmatrix} 1 & 0.51 & 0.81 & 0.82 \\ 0.51 & 1 & 0.77 & 0.79 \\ 0.81 & 0.77 & 1 & 0.49 \\ 0.81 & 0.79 & 0.49 & 1 \end{pmatrix},$$

$$Q_{i,j,1} = \begin{pmatrix} 0.9 & 0.1 & 0 & 0 \\ 0.105 & 0.895 & 0 & 0 \\ * & * & * & * \\ * & * & * & * \end{pmatrix}, Q_{i,j,2} = \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ 0 & 0 & 0.75 & 0.25 \\ 0 & 0 & 0.26 & 0.74 \end{pmatrix}.$$

Figure 5.3.12. Complete 3-dimensional phase portrait for an asymmetric case.

The results of this section provide some important intuition for thinking about mutations that affect UG, and in particular, how selection and mutation might improve UG over time. Imagine that we have a population in which everyone has the same UG, say, U_1. In the model, this corresponds to a population state on the top edge of the simplex. The introduction of a mutant with a modified UG, say U_2, corresponds to a slight perturbation away from the edge. Whether this mutation dies out or propagates determines whether or not it can be incorporated into the population.

Suppose first that U_2 is essentially the same as U_1, that is, the mutation is in some non-critical area and universal grammar is barely affected. This situation appears in Figure 5.3.10. Homogeneous populations are unstable, and the introduction of a mutant is likely to lead to a mixed population. From there, factors not included in this model may guide the population to choose one UG or the other, but the important conclusion is that modifications to UG that are highly compatible with existing grammars do not die out immediately; in fact, they may persist indefinitely.

Now suppose that U_2 is moderately similar to U_1, as depicted in Figure 5.3.9. In this case, the outcome depends on the learning algorithm associated with U_1 and the linguistic environment. Grammar acquisition may be sufficiently accurate that the population is stable against invasion by U_2, as in the top two pictures. If it is error prone as in the bottom picture, the population could settle into a mixed state. Thus, changes to UG introducing minor incompatibilities may or may not die out immediately.

The final case is that of a drastic change, in which U_2 is significantly different from U_1. This situation, shown in Figure 5.3.11, has the property that homogeneous populations are always stable against invasion, so mutations introducing large incompatibilities die out immediately.

The lesson here is that an innovative form of communication is useless unless it is compatible with the existing population. A mutant with a significantly better but incompatible universal grammar would be unable to realize any benefit from it because there would be no one to talk to, and the innovation would die out.

With these comments in mind, we may describe the continuing evolution of UG in terms of an annealing process, with learning error analogous to temperature. Sometimes, the learning error is relatively high and allows moderately incompatible mutations to survive with high probability. This would allow the population to explore the fitness landscape fairly quickly. At other times, learning might become more reliable either through improvements to the internal algorithm for grammar acquisition or through changes to the linguistic environment. Under these circumstances, moderately incompatible mutations are weeded out because the benefit from communicative compatibility would in general out weigh the benefit of further large changes, but UG could still change through incremental innovations.

5.4. Partial results for the general case

The goal of this section is to return to the original form of the language dynamical equation with multiple universal grammars from Section 5.2 and obtain some theoretical results without making any simplifying assumptions about the parameter matrices B and Q. In its most general form, this model has the property that the learning algorithm does not explicitly appear in the time derivative of the size of the sub-population with a given U_K, shown in (5.2.4). This observation motivates a series of calculations for the case of two universal grammars with two grammars each that yields a set of sufficient conditions for exclusion to exist, that is, the sets of population states in which everyone has the same UG are attracting. The point of developing this result is that the sufficient conditions consist of six fairly simple inequalities that may be easily interpreted in a number of special cases, thus building mathematical intuition for when competitive exclusion is inevitable.

An *attracting set* [26] is a closed invariant subset of the phase space surrounded by a neighborhood in which every trajectory tends to the set in forward time. As simple calculation using (5.2.4) shows that $\dot{y}_K = 1$ if $y_K = 0$, so for each K the set of points where $y_K = 1$ is closed and invariant. The argument that it is attracting is based on determining some geometric properties of a surface in the simplex. This surface divides the simplex into three regions, one in which all populations tend toward the top edge where U_1 takes over, one in which all populations tend toward the bottom edge where U_2 takes over, and an intermediate region. To illustrate the argument, we will first analyze a case with highly symmetric parameter settings, then generalize the argument to the case of asymmetric parameter settings.

5.4.1. Illustration of the null-cline argument in the case of permutation symmetry. To illustrate the argument, consider the case where all the grammars overlap equally, so

$$B = \begin{pmatrix} 1 & a & a & a \\ a & 1 & a & a \\ a & a & 1 & a \\ a & a & a & 1 \end{pmatrix}. \tag{5.4.1}$$

This case happens to be a specialization of the case studied in Section 5.3, so we can already determine a great deal about it. However, it also provides an illustration of the null-cline argument which is needed for the general case in which there are no simple expressions for the various fixed points.

For convenience, define $U(i) = K$ such that U_K generates G_j. With that notation, the fitness of G_j simplifies considerably into

$$\begin{aligned} F_j &= \sum_{i=1}^{n} B_{i,j} x_{i,U(i)} \\ &= x_{j,U(j)} + a \sum_{i \neq j} x_{i,U(i)} \\ &= x_{j,U(j)} + a\left(1 - x_{j,U(j)}\right) \\ &= a + (1-a) x_{j,U(j)}. \end{aligned}$$

Likewise,

$$\begin{aligned} \phi &= \sum_{K=1}^{N} a y_K + (1-a) \sum_{j=1}^{n} x_{j,U(j)} x_{j,K} \\ &= a + (1-a) \sum_{j=1}^{n} \left(x_{j,U(j)} \sum_{K=1}^{N} x_{j,K} \right) \\ &= a + (1-a) \sum_{j=1}^{n} x_{j,U(j)}^2. \end{aligned}$$

There are two universal grammars, so we have two variables of interest, $y_1 = x_{1,1} + x_{2,1}$ and $y_2 = x_{3,2} + x_{4,2}$. Since $y_1 + y_2 = 1$, we need only analyze the behavior of y_1. The following proposition, describes how the limiting behavior of y_1 is largely determined by the initial population state.

Proposition 5.4.1. *The simplex contains two trapping regions which are independent of the Q matrix: Trajectories for which $y_1(0) > 2/3$ tend to $y_1 = 1$, and trajectories for which*

$y_1(0) < 1/3$ *tend to* $y_1 = 0$*. In the region in between, the* Q *matrix influences whether* y_1 *approaches* 1 *or* 0*.*

Proof. As in Section 5.3.2, we change from simplex coordinates to box coordinates. Observe that $y_1 = z$ in these new coordinates, from which it follows that

$$\dot{z} = \dot{y}_1 = \frac{1}{2}(1-a)(1-z)z(-1-s^2(1-z)+(2+r^2)z). \tag{5.4.2}$$

To find the Q-independent trapping regions, we first look for the z null-clines. These are the sets of points for which $\dot{z} = 0$. From (5.4.2) it is clear that $\dot{z} = 0$ if and only if $z = 0$, $z = 1$, or $z = (1+s^2)/(2+s^2+r^2)$. The first two cases are the upper and lower edges of the simplex, and the third is a surface near $z = 1/2$. See Figure 5.4.1. Thus, $\dot{z}$ is of one sign above the surface and the opposite sign below. Looking at the vertical line given by $\{r = 0, s = 0\}$, we have $\dot{z} = -\frac{1}{2}z(-1+z)(-1+2z)$ which is positive for $z > 1/2$ and negative for $z < 1/2$.

Therefore, the overall picture is that the simplex decomposes into upper and lower trapping regions and a boundary region in the middle. If a trajectory starts above the topmost point of the z null-cline, then $\dot{y}_1 > 0$, which means y_1 will increase over time until it reaches $y_1 = 1$. Likewise, any trajectory that starts below the bottommost point of the surface will continue downward until $y_1 = 0$.

To find the topmost and bottommost points, observe that the surface is saddle shaped, so the extrema will appear on the boundaries. Define the function

$$h(r,s) = \frac{1+s^2}{2+s^2+r^2},$$

whose graph is the surface in question. Looking on the faces of the simplex given by $s = \pm 1$, we have $h(r, \pm 1) = 2/(3+r^2)$ which has a maximum at $h(0, \pm 1) = 2/3$ and minima on the edges at $h(\pm 1, \pm 1) = 1/2$. Likewise, looking on the surfaces given by $r = \pm 1$, we have $h(\pm 1, s) = 1 - 2/(3+s^2)$ which has a minimum at $h(\pm 1, 0) = 1/3$. Furthermore, it has maxima on the edges at $h(\pm 1, \pm 1) = 1/2$. Therefore, if either universal grammar holds a 2/3 majority of the population, it will eventually take over regardless of the Q matrix. □

In the boundary region near the z null-cline, many orbits obey the simple rule that if they start above the surface, they approach $y_1 = 1$ and if they start below, they approach $y_1 = 0$. However, orbits may pass through the surface horizontally, thereby starting above it but converging to $y_1 = 0$ or starting below it but converging to $y_1 = 1$. For example, Figure 5.4.2 shows the values of y_1 and y_2 starting from a point just above the z null-cline for which $y_1 \to 1$. However, a nearby initial condition produces the trajectories in Figure 5.4.3, where y_1 passes horizontally through the z null-cline and turns downward.

The actual surface dividing orbits that go to $y_1 = 1$ from those that go to $y_1 = 0$ appears to be the stable manifold of the saddle point $C0$ in the middle of the simplex, and this manifold depends on Q.

The proposition implies that in this case, the learning algorithms employed by the two universal grammars and specified by Q are largely irrelevant to determining which universal

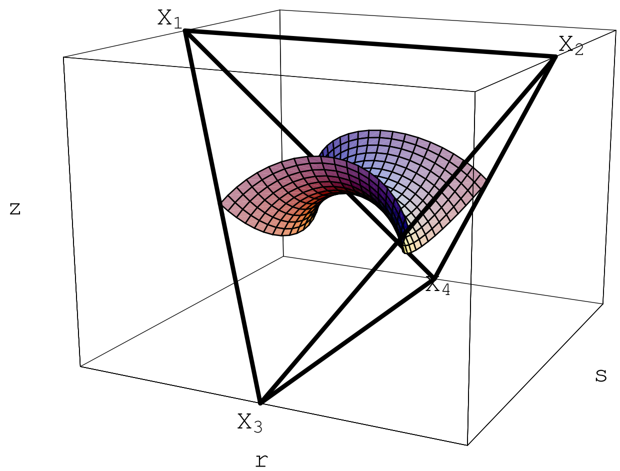
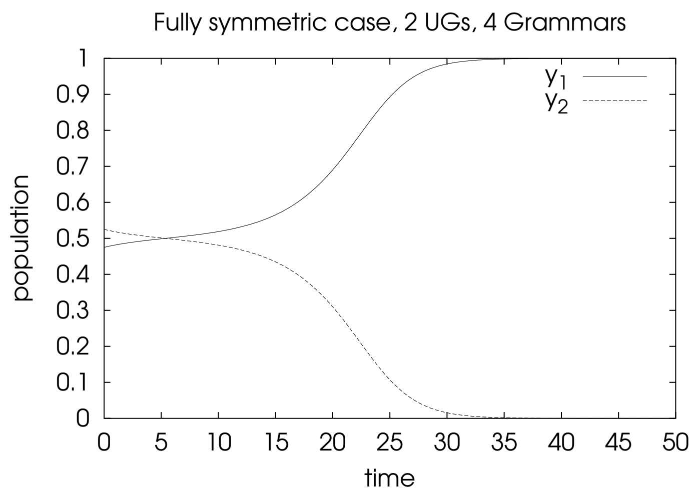


Figure 5.4.1. The upended simplex and z null-cline.



$$a = 1/10, Q_{i,j,1} = \begin{pmatrix} 0.7 & 0.3 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, Q_{i,j,2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.7 & 0.3 \\ 0 & 0 & 0.3 & 0.7 \end{pmatrix}$$

Figure 5.4.2. Trajectories starting from $x_{1,1} = 57/160 = 0.35625, x_{2,1} = 19/160 = 0.11875, x_{3,2} = x_{4,2} = 21/80 = 0.2625$ which lies just above the z null-cline.

grammar takes over the population. As long as a certain majority of the population uses one universal grammar, only the initial population state matters.

5.4.2. Null-cline argument in the general case. In this subsection, we extend the null-cline argument of Section 5.4.1 to the case of fully general B and Q matrices, assuming

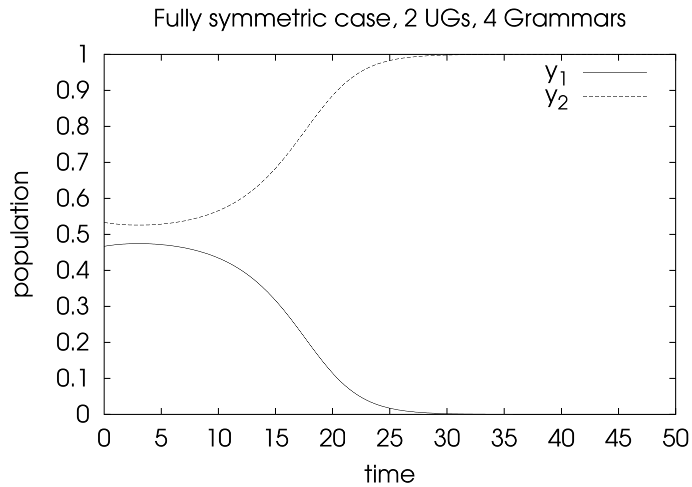


Figure 5.4.3. Trajectories starting from $x_{1,1} = 7/20 = 0.35, x_{2,1} = 7/60 = 0.11\bar{6}, x_{3,2} = x_{4,2} = 4/15 = 0.2\bar{6}$. Here, the trajectory passes through the z null-cline, and y_1 increases, reaches a maximum, then turns downward and tends to 0. The parameters are the same as in Figure 5.4.2.

that there are two universal grammars, each of which admits two grammars. We first find sufficient conditions under which the null-cline does not intersect the top and bottom edges of the simplex. This guarantees that there are regions which lie completely above or below it. Second, we find a condition which implies that orbits above the surface move upward, and those below it move downward, thereby assuring that the top and bottom edges are attracting sets.

We will allow the B matrix to be completely general, lifting even the requirement that it be symmetric:

$$B = \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{pmatrix}.$$

This generality allows the results that follow to apply to variations of this model that might incorporate additional information such as ambiguity into the measurement of payoff. In what follows, more concise expressions result if the following parameters are used instead of the entries of B:

$$\alpha_0 = \frac{1}{2}(b_{11} + b_{12} + b_{21} + b_{22}) \qquad \alpha_1 = \frac{1}{2}(b_{11} - b_{12} - b_{21} + b_{22})$$
$$\alpha_2 = \frac{1}{2}(b_{11} + b_{12} - b_{21} - b_{22}) \qquad \alpha_3 = \frac{1}{2}(b_{11} - b_{12} + b_{21} - b_{22})$$
$$\beta_0 = \frac{1}{2}(b_{13} + b_{14} + b_{23} + b_{24}) \qquad \beta_1 = \frac{1}{2}(b_{13} - b_{14} - b_{23} + b_{24})$$
$$\beta_2 = \frac{1}{2}(b_{13} + b_{14} - b_{23} - b_{24}) \qquad \beta_3 = \frac{1}{2}(b_{13} - b_{14} + b_{23} - b_{24})$$
$$\gamma_0 = \frac{1}{2}(b_{31} + b_{32} + b_{41} + b_{42}) \qquad \gamma_1 = \frac{1}{2}(b_{31} - b_{32} - b_{41} + b_{42})$$
$$\gamma_2 = \frac{1}{2}(b_{31} + b_{32} - b_{41} - b_{42}) \qquad \gamma_3 = \frac{1}{2}(b_{31} - b_{32} + b_{41} - b_{42})$$
$$\delta_0 = \frac{1}{2}(b_{33} + b_{34} + b_{43} + b_{44}) \qquad \delta_1 = \frac{1}{2}(b_{33} - b_{34} - b_{43} + b_{44})$$
$$\delta_2 = \frac{1}{2}(b_{33} + b_{34} - b_{43} - b_{44}) \qquad \delta_3 = \frac{1}{2}(b_{33} - b_{34} + b_{43} - b_{44})$$

We will work in box coordinates again, as defined in (5.3.3). After some simplification,

$$\dot{z} = \frac{1}{4}(-1 + z)zg(r, s, z), \tag{5.4.3}$$

where

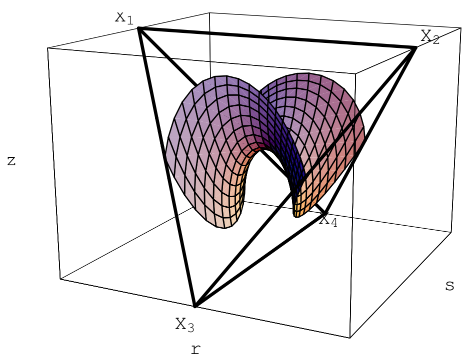
$$\begin{aligned} g(r, s, z) =2\Big(& - \beta_0 + \delta_0 + r\beta_2 + s(\beta_3 - \delta_2 - \delta_3) - z(\alpha_0 - \beta_0 - \gamma_0 + \delta_0) \\ & - rs\beta_1 + rz(\alpha_2 + \alpha_3 - \beta_2 - \gamma_3) - sz(\beta_3 + \gamma_2 - \delta_2 - \delta_3) \\ & + s^2\delta_1 + rsz(\beta_1 + \gamma_1) - r^2 z\alpha_1 - s^2 z\delta_1\Big) \end{aligned} \tag{5.4.4}$$

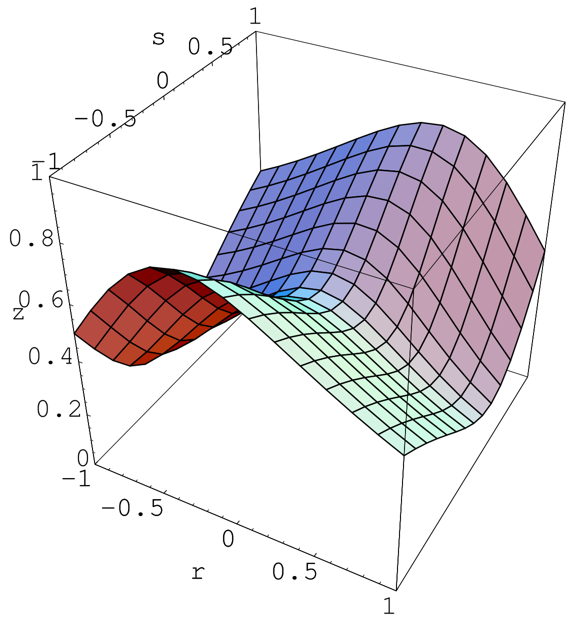
The form of $\dot{z}$ in this general case is similar to the form (5.3.5) in the symmetric case, and as before there are three z null-clines: the top ($z = 1$), the bottom ($z = 0$), and the surface determined by $g(r, s, z) = 0$. This surface will be called the interior z null-cline. The goal of this section is to determine sufficient conditions on B such that the interior z null-cline creates trapping regions around the top and bottom edges of the simplex.

5.4.2.1. *Step 1: The non-intersection constraints.* The first condition is that the interior z null-cline must not touch the top and bottom. This condition implies that there is some space between the vertical extrema of the surface and the top and bottom edges of the simplex. (See Figure 5.4.4.)

Proposition 5.4.2. *Assume that $\alpha_1 \neq 0$ and $\delta_1 \neq 0$. Suppose further that the following expressions are all positive:*

$$\begin{aligned} \nu_1 &= 4\delta_1(\delta_0 - \beta_0 + \beta_2) - (\beta_1 - \beta_3 + \delta_2 + \delta_3)^2 \\ \nu_2 &= 4\delta_1(\delta_0 - \beta_0 - \beta_2) - (\beta_1 + \beta_3 - \delta_2 - \delta_3)^2 \\ \nu_3 &= 4\alpha_1(\alpha_0 - \gamma_0 + \gamma_2) - (\alpha_2 + \alpha_3 + \gamma_1 - \gamma_3)^2 \\ \nu_4 &= 4\alpha_1(\alpha_0 - \gamma_0 - \gamma_2) - (\alpha_2 + \alpha_3 - \gamma_1 - \gamma_3)^2 \end{aligned} \tag{5.4.5}$$





$$B = \begin{pmatrix} 1 & \frac{1}{4} & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & 1 & \frac{1}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{1}{4} & 1 & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} & \frac{1}{4} & 1 \end{pmatrix}$$

Figure 5.4.4. The interior z null-cline in simplex coordinates (left) and box coordinates (right). The asymmetric B matrix used to generate these pictures is as shown.

and also that

(5.4.6) $$\beta_0 - \delta_0 > 0 \text{ and } \gamma_0 - \alpha_0 > 0.$$

Then, the interior z null-cline lies strictly between the top and bottom of the simplex at a strictly positive distance from each.

Proof. The mathematical formulation of the conclusion in box coordinates is that if $-1 \leq r \leq 1$ and $-1 \leq s \leq 1$, then $g(r, s, 0) \neq 0$ and $g(r, s, 1) \neq 0$. We proceed by proving that g is of one sign on the bottom plane $z = 0$, and also of one sign on the top plane $z = 1$. The technical assumptions that α_1 and δ_1 are nonzero eliminate some degenerate cases that would cause division by zero in what follows.

For the bottom, we are interested in $g(r, s, 0)$, which happens to be a quadratic form in r and s, so the equation $g(r, s, 0) = 0$ must define a conic section in the plane $z = 0$. To classify it, we complete the square in r and s and change variables to ρ and σ so as to put it in a standard form:

$$r = \frac{2\beta_2\delta_1 + \beta_1(\beta_3 - \delta_2 - \delta_3 - 2\delta_1\rho)}{\beta_1^2}.$$

$$s = \frac{\beta_2}{\beta_1} - \rho + \sigma.$$

With these new variables, the equation $g(r, s, 0) = 0$ becomes

$$-\frac{2(\beta_0\beta_1^2 - \beta_1^2\delta_0 - \beta_2^2\delta_1 + \beta_1\beta_2(-\beta_3 + \delta_2 + \delta_3))}{\beta_1^2} - 2\delta_1\rho^2 + 2\delta_1\sigma^2 = 0,$$

which is the form of a hyperbola in ρ and σ. So if we want to specify that the interior z null-cline does not touch the bottom, it is sufficient to require that the hyperbola specified by $g(r, s, 0) = 0$ lies outside the square given by $-1 \leq r \leq 1$ and $-1 \leq s \leq 1$. (See Figure 5.4.5.) That constraint is equivalent to requiring the expression $g(r, s, 0)$ to be of

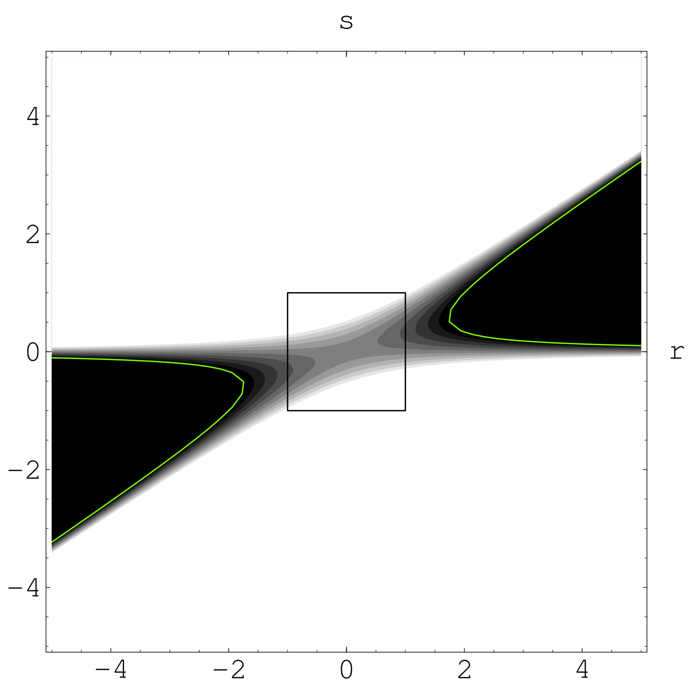


Figure 5.4.5. Contour plot of $g(r, s, 0)$. Darker values are negative, lighter values are positive. The light hyperbola is $g(r, s, 0) = 0$, that is, the curve where the z null-cline intersects the plane $z = 0$. The square is the bottom face of the phase space. See Figure 5.4.4 for the particular B used in this illustration. This is where we have to think outside the box.

one sign on the sides of the square. To avoid having two separate cases ($g > 0$ or $g < 0$), we transform the constraint by dividing $g(r, s, 0)$ by the coefficient of s^2 and requiring the resulting expression $g_2(r, s)$ to be positive on the sides of the square:

$$g_2(r, s) = s^2 + \frac{-\beta_0 + r\beta_2 + \delta_0}{\delta_1} - \frac{s(r\beta_1 - \beta_3 + \delta_2 + \delta_3)}{\delta_1}. \tag{5.4.7}$$

(See Figure 5.4.6.)

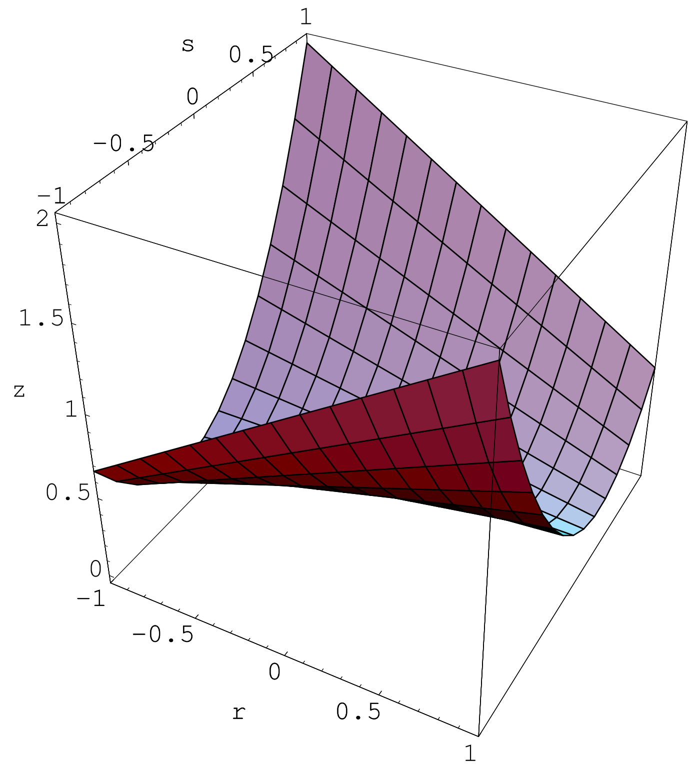


Figure 5.4.6. Plot of $g_2(r,s)$, which differs by a constant factor from $g(r,s,0)$, using the same B as in Figure 5.4.4. The region shown is the phase space in box coordinates. Observe that the surface intersects with two faces of the phase space in parabolas, and with the other in lines.

Note that on the sides where $r = \pm 1$, the expressions $g(\pm 1, s)$ are quadratic functions of s. So, to guarantee that $g_2(r,s) > 0$ on these two sides, it suffices to require that the minima of $g_2(\pm 1, s)$ be positive. The minimum of a general quadratic function $x^2 + ax + b$ is $b - a^2/4$, so the exact constraints are

$$\min_s g_2(1,s) = -\frac{-4(-\beta_0 + \beta_2 + \delta_0)\delta_1 + (\beta_1 - \beta_3 + \delta_2 + \delta_3)^2}{4\delta_1^2} > 0,$$
$$\min_s g_2(-1,s) = -\frac{4(\beta_0 + \beta_2 - \delta_0)\delta_1 + (\beta_1 + \beta_3 - \delta_2 - \delta_3)^2}{4\delta_1^2} > 0.$$

Both denominators are square, so only the numerators matter in satisfying the inequalities. We therefore simplify the constraints to the first two inequalities in the statement of the proposition, namely $\nu_1 > 0$ and $\nu_2 > 0$.

Observe that if these constraints are satisfied, then $g_2(r,s) > 0$ on all four corners of the square. With that observation, the sides where $s = \pm 1$ are easy to check, as $g_2(r, \pm 1)$ is a linear function of r, and it is therefore enough to require that $g_2(r,s) > 0$ on the corners.

In summary, if $\nu_1 > 0$ and $\nu_2 > 0$, then g is of one sign on all four sides of the bottom square in box coordinates, and therefore, the z null-cline does not intersect the bottom of the simplex.

The constraint that the interior z null-cline cannot intersect with the top of the simplex can be enforced by imposing a second set of inequalities similar to those discovered above. Again, the equation for where the null-cline intersects $z = 1$ is $g(r, s, 1) = 0$ which defines a hyperbola in the plane $z = 1$ in terms of r and s. To specify that the null-cline does not touch the top edge of the simplex, it suffices to require that this hyperbola lie outside the square in box coordinates given by $-1 \le r \le 1$ and $-1 \le s \le 1$. As before, we ensure this by requiring $g(r, s, 1)$ to be of one sign on all four sides of the square. Equivalently, we define $g_3(r, s)$ to be $g(r, s, 1)$ divided by the coefficient of r^2, and require $g_3(r, s)$ to be positive on all four sides of the square. The expression for g_3 is

$$g_3(r, s) = r^2 + \frac{\alpha_0 - \gamma_0 + s\gamma_2}{\alpha_1} - \frac{r(\alpha_2 + \alpha_3 + s\gamma_1 - \gamma_3)}{\alpha_1}. \tag{5.4.8}$$

Furthermore, $g_3(r, \pm 1)$ are monic quadratic functions of r, so it suffices to require that their minima be positive, which yields

$$\min_r g_3(r, 1) = -\frac{-4\alpha_1(\alpha_0 - \gamma_0 + \gamma_2) + (\alpha_2 + \alpha_3 + \gamma_1 - \gamma_3)^2}{4\alpha_1^2} > 0,$$
$$\min_r g_3(r, -1) = -\frac{4\alpha_1(-\alpha_0 + \gamma_0 + \gamma_2) + (\alpha_2 + \alpha_3 - \gamma_1 - \gamma_3)^2}{4\alpha_1^2} > 0.$$

As before, the denominators are all square, so only the numerators matter, and the constraints reduce to $\nu_3 > 0$ and $\nu_4 > 0$. These imply that g is of one sign on the top of the phase space in box coordinates, and that the z null-cline does not intersect with the top of the simplex.

The final two constraints in the statement of the proposition are there to ensure that the interior null-cline lies inside the simplex rather than completely above or below it, and are derived as follows. Choose $\bar{z}$ such that $g(0, 0, \bar{z}) = 0$, that is, the point at which the null-cline intersects the vertical line given by $r = 0$ and $s = 0$:

$$\bar{z} = \frac{\delta_0 - \beta_0}{\alpha_0 - \gamma_0 + \delta_0 - \beta_0}. \tag{5.4.9}$$

A short calculation proves that the constraints $\delta_0 - \beta_0 > 0$ and $\alpha_0 - \gamma_0 > 0$ imply $0 < \bar{z} < 1$, which guarantees that the null-cline lies completely inside the simplex.

It turns out that $g(r, s, z) = 0$ can actually be solved in terms of z, and the resulting solution $z = h(r, s)$ is the quotient of two polynomials in r and s. Under the constraints derived in this proposition, h must be bounded for $-1 \le r \le 1$ and $-1 \le s \le 1$, which means its denominator never vanishes. Therefore, h is continuous, and since the region of interest for r and s is a closed square, h actually takes on its extrema. It follows that there is a strictly positive distance between the null-cline and the top and bottom of the simplex. □

It is worth mentioning why we constrain g_2 and g_3 to be positive rather than negative. This choice is motivated by the fully symmetric case where

$$B = \begin{pmatrix} c & a & a & a \\ a & c & a & a \\ a & a & c & a \\ a & a & a & c \end{pmatrix},$$

in which $\nu_1 = \nu_2 = \nu_3 = \nu_4 = 4(c-a)^2$ which is always positive. Furthermore, these four constraints are generally consistent with diagonally dominant settings for B, which match the intuition that in general, people should have the greatest probability of communicating properly when they both use exactly the same grammar.

5.4.2.2. *Step 2: Direction of the vector field.*

Proposition 5.4.3. *Assume that the interior z null-cline is strictly between the top and bottom edges of the simplex, and that $\delta_0 - \beta_0 > 0$ and $\alpha_0 - \gamma_0 > 0$. Then there are trapping regions above and below the null-cline.*

Proof. We will work in box coordinates and show that $\dot{z} > 0$ above the interior z null-cline, and $\dot{z} < 0$ below it. This claim implies that orbits that pass above the uppermost point on the null-cline continue to rise, and those that pass below the lowermost point continue to fall, thereby establishing the existence of the two trapping regions.

The null-clines are by definition the set of points where $\dot{z} = 0$, so in regions between them, $\dot{z}$ is of one sign. It therefore suffices to show that for some point above the interior null-cline, $\dot{z} > 0$, and for some point below it, $\dot{z} < 0$. Consider the vertical line given by $r = 0$ and $s = 0$, as illustrated in Figure 5.4.7. Along this line,

$$\dot{z}|_{r=0,s=0} = \frac{1}{2}(-1+z)z(-\beta_0 + z(-\alpha_0 + \beta_0 + \gamma_0 - \delta_0) + \delta_0).$$

That is, $\dot{z}$ is a cubic function $f(z)$ along this vertical line, as in Figure 5.4.8. We need only require that $f'(0) < 0$ to ensure that $\dot{z}$ is negative below the null-cline and positive above it, which is equivalent to the inequality

$$f'(0) = \frac{\beta_0 - \delta_0}{2} < 0. \tag{5.4.10}$$

Equivalently, we may require that $f'(1) < 0$, which yields the inequality

$$f'(1) = \frac{\gamma_0 - \alpha_0}{2} < 0. \tag{5.4.11}$$

Both of these inequalities follow immediately from the hypotheses. □

5.4.3. Discussion. What makes these two propositions possible is the fact that Q is row-stochastic, so that it disappears in $\dot{y}_K$ in (5.2.4). The same would happen if Q depended on t or x. The trapping regions described in Propositions 5.4.2 and 5.4.3 are also independent of Q. Thus, for any collection of grammars that satisfies the hypotheses of these propositions, as long as one of the universal grammars has a sufficiently large majority of the population, it will take over no matter what the learning algorithm, static or dynamic. Although these

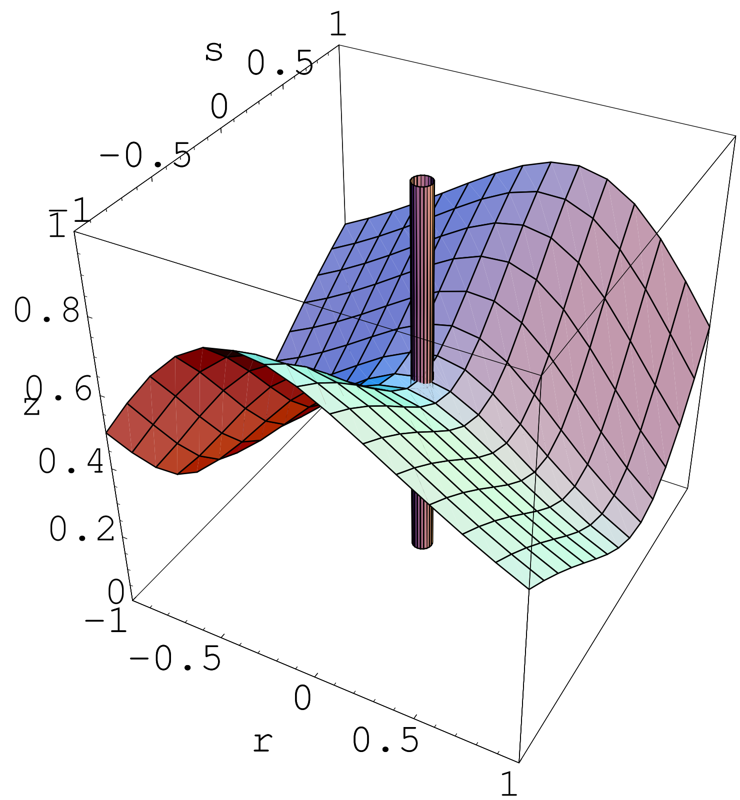
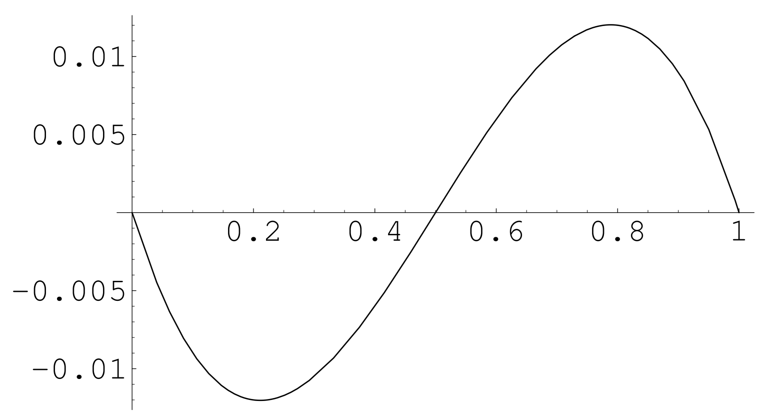


Figure 5.4.7. The z null-cline in box coordinates with the line $r = 0, s = 0$ indicated by a bar. The B matrix used for this picture is the same as in Figure 5.4.4



$$f(z) = \dot{z}|_{r=0,s=0}$$

Figure 5.4.8. The vertical component of the vector field along the central vertical line in Figure 5.4.7

propositions do not rule out the possibility of stable coexistence, they do provide conditions under which homogeneous populations are stable against invasion by the other UG.

The explicit form of these results provides a way to quickly examine any number of parameter settings so as to develop mathematical intuition for what properties of the B matrix lead to stable UGs. For example, consider the question of two very different UGs, where G_1 and G_2 do not communicate well with G_3 and G_4. The payoff matrix for such a situation might look like this:

$$B_{\text{diff}} = \begin{pmatrix} c & a & \varepsilon & \varepsilon \\ a & c & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & c & a \\ \varepsilon & \varepsilon & a & c \end{pmatrix}. \tag{5.4.12}$$

where c is relatively large, ε is small, and a is in between. For a picture, see Figure 5.4.9. The constraints simplify greatly in this case:

$$\nu_1 = \nu_2 = \nu_3 = \nu_4 = 4(c-a)(c+a-2\varepsilon),$$

$$\alpha_0 - \gamma_0 = \delta_0 - \beta_0 = a + c - 2\varepsilon.$$

Clearly, if ε is small enough, then all six constraints are positive. Therefore, the two UGs are stable against invasion by each other no matter what their learning processes are.

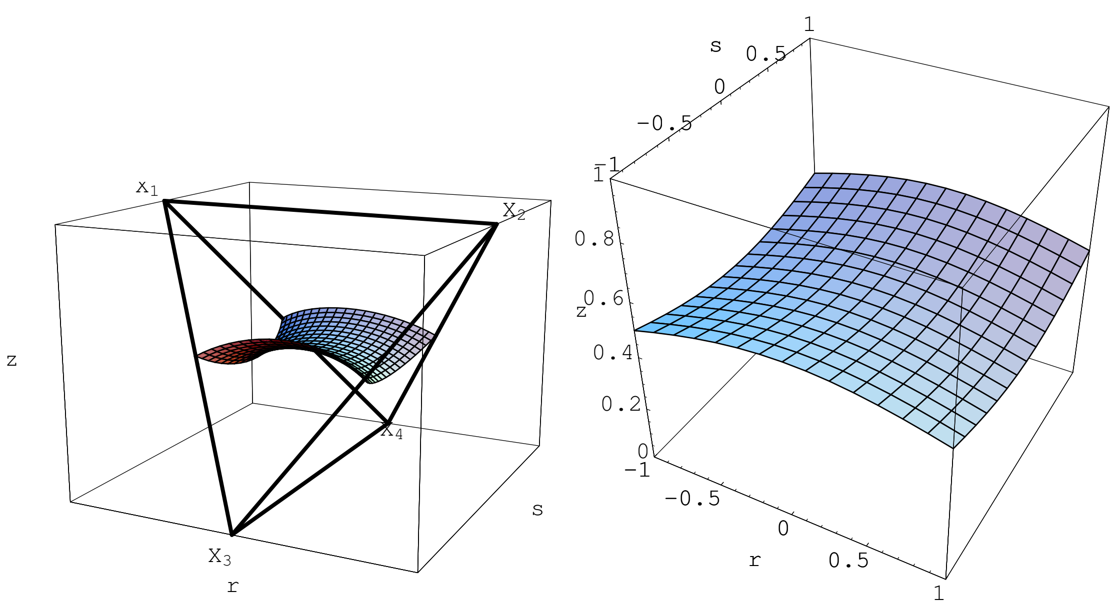


Figure 5.4.9. Null-cline for the case of two very different UGs, using the payoff matrix B_{diff} with $c = 1$, $a = 1/2$, and $\varepsilon = 1/8$. Left: Phase space in simplex coordinates. Right: Phase space in box coordinates.

On the other hand, consider the case of two very similar UGs, where $G_1 \approx G_3$ and $G_2 \approx G_4$. The payoff matrix for this example might look like this:

$$B_{\mathrm{sim}} = \begin{pmatrix} c & a & (1-\varepsilon)c & (1-\varepsilon)a \\ a & c & (1-\varepsilon)a & (1-\varepsilon)c \\ (1-\varepsilon)c & (1-\varepsilon)a & c & a \\ (1-\varepsilon)a & (1-\varepsilon)c & a & c \end{pmatrix}. \tag{5.4.13}$$

The constraints simplify to

$$\nu_1 = \nu_2 = \nu_3 = \nu_4 = -(c-a)^2 - 2(a^2 + 2ac - 3c^2)\varepsilon - (a-c)^2\varepsilon^2,$$

and

$$\alpha_0 - \gamma_0 = \delta_0 - \beta_0 = (a+c)\varepsilon.$$

If ε is small enough, then ν_1, ν_2, ν_3 and ν_4 are dominated by $-(c-a)^2$ which is negative, implying that the null-cline might intersect with the top and bottom of the phase space as illustrated in Figure 5.4.10. The propositions do not apply in this case, so it is possible that one UG might be able to invade the other, and the learning process is critical to understanding the long-term behavior of the system.

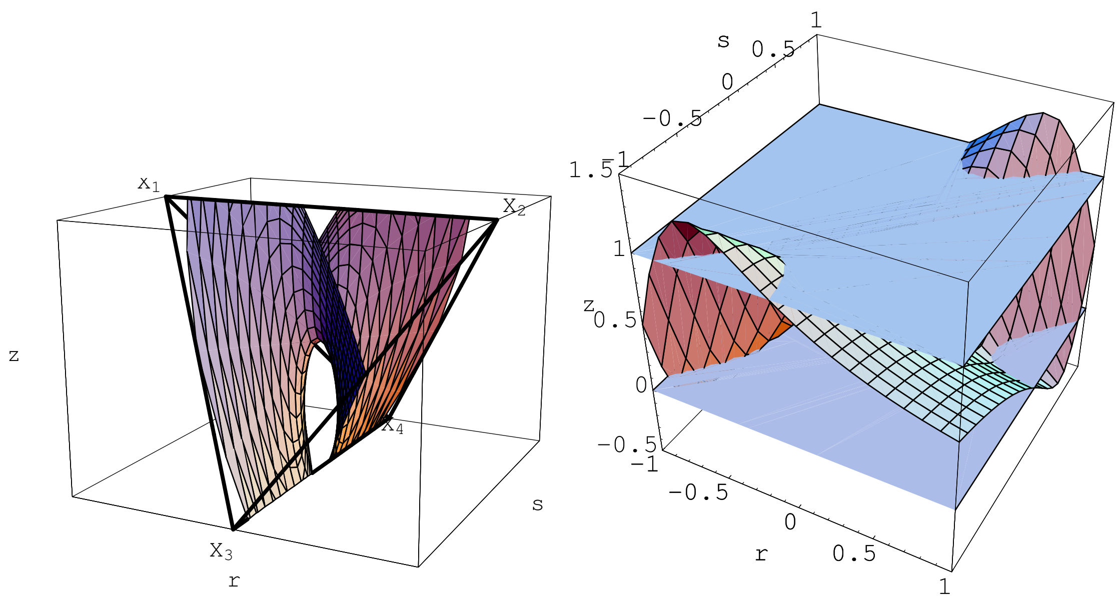
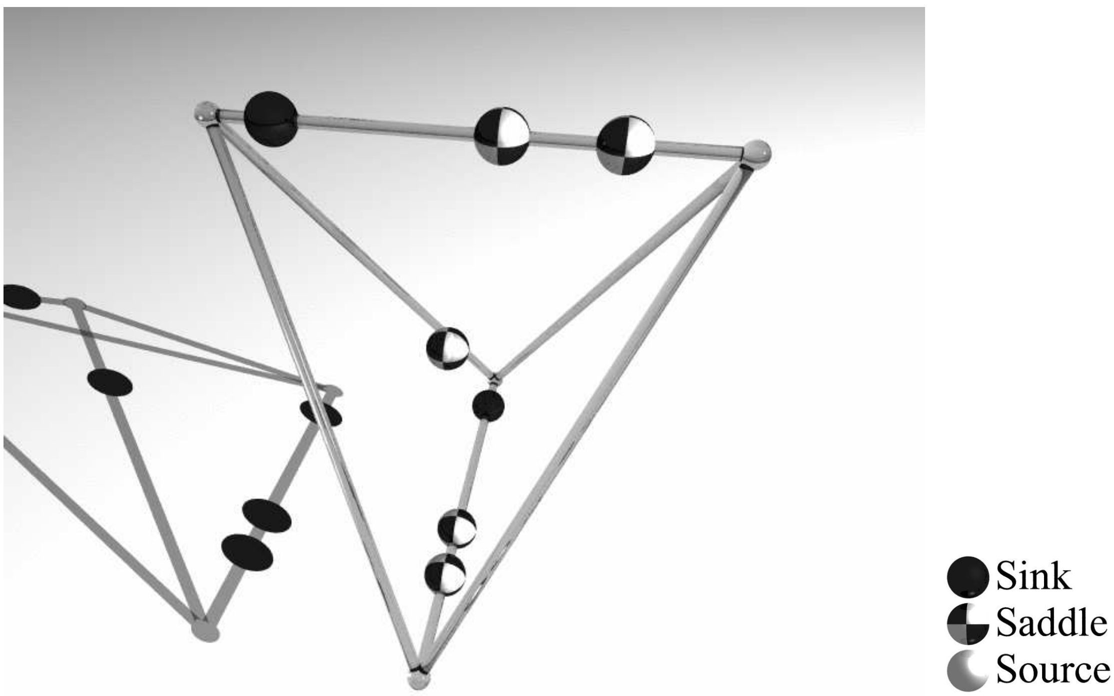


Figure 5.4.10. Null-cline for the case of two very similar UGs, using the payoff matrix B_{sim} with $c = 1$, $a = 1/2$, and $\varepsilon = 1/32$. Left: Phase space in simplex coordinates. Right: Phase space in box coordinates, with planes $z = 0$ and $z = 1$ indicated. Note that the null-cline intersects with these planes within the phase space.

The case of similar grammars leads to a remarkable situation that might be called *accidental stability,* depicted in Figure 5.4.11. This is a case of two similar UGs where $G_1 \approx G_3$ and $G_2 \approx G_4$, so it is no surprise that the two UGs can invade one another. What is surprising is the mechanism. Consider an initial population whose members all have U_1. These states are all on the top edge of the simplex and remain on that line unless subject to an external perturbation. Such a population will tend to one of the two fixed points on the top edge, one of which is dominated by G_1 and the other by G_2. The fixed point dominated by G_1 is a stable sink, and if U_2 tries to invade that population, it will fail. However, the fixed point dominated by G_2 is a saddle, and if U_2 tries to invade that population, the invasion succeeds, U_2 takes over completely, and the population tends to the sink on the bottom edge of the simplex dominated by G_4. The initial state of the all-U_1 population determines whether a later invasion by U_2 succeeds or not, and that initial state is essentially random. Hence, this instance of the language equation is sensitive to historical accidents.



$$B = \begin{pmatrix} 1 & 0.2 & 0.99 & 0.1 \\ 0.2 & 1 & 0.1 & 0.99 \\ 0.99 & 0.1 & 1 & 0.2 \\ 0.1 & 0.99 & 0.2 & 1 \end{pmatrix} \quad Q = \left[\begin{pmatrix} 0.908 & 0.092 & 0 & 0 \\ 0.130 & 0.870 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.858 & 0.142 \\ 0 & 0 & 0.092 & 0.908 \end{pmatrix} \right]$$

Figure 5.4.11. Accidental stability.

5.5. Conclusion

The analysis in Section 5.3 shows that the language dynamical equation with multiple universal grammars can exhibit dominance, competitive exclusion, and coexistence in the case of two UGs with two grammars each. Although the results are for symmetric parameter settings, the phase portraits appear to be structurally stable, so all three behaviors should persist when the parameters are perturbed. Stable coexistence of multiple universal grammars appears to be possible for a range of parameter values. Exclusion is the only option if the two UGs are sufficiently different. For a range of parameter values, homogeneous populations are stable against invasion. This fact has significant consequences for the evolution of UG: The benefits of communicating with the rest of the population limit the population to innovations that are fairly compatible with the existing UG. Other beneficial mutations are likely to die out before their benefits can be realized. Sufficient conditions for the exclusion of incompatible mutations in general can also be found, as in Section 5.4.

This research could be extended in a number of directions. The question of when can an innovative mutation survive suggests that a stochastic model of a finite population might be enlightening. The results of Section 5.4 are independent of the learning algorithm, and give no indication of how acquisition might change over time. Thus, it would be informative to study cases of the model for two UGs that differ only in their learning algorithm. It would also be interesting to add terms to the fitness function for trade-offs. For example, one learning algorithm might be very precise but take significantly longer, thereby penalizing its carriers as well as giving them the benefits of precise learning. Another might have few hypotheses and therefore be fast and precise, but only at the cost of encoding more information genetically. Also, the linguistic environment could be modeled in more detail, including features such as noisy data and locality.

Chapter 6

Discussion, Conclusion and Future Possibilities

Contents

6.1. Summary

6.1.1. Modeling language change. We started with a model of language dynamics that combines game dynamics, derived from the simplifying assumption that survival is based on the ability to communicate, with a learning process. The resulting language dynamical equation can be instantiated in any number of dimensions and at any level of complexity. To begin making progress, we examined highly symmetric parameter settings, and low dimensional cases, both of which can mimic observed patterns of language change, but with very different mechanisms.

In the highly symmetric case of Chapter 2, all population states tend to some stable equilibrium, and the learning accuracy determines whether it is possible to have an equilibrium where all grammars are present, or an equilibrium where one grammar dominates.

The model can illustrate catastrophes that occur when languages are brought into contact if the learning parameters are allowed to be non-constant on long time scales: Contact can be simulated by suddenly reducing the learning accuracy to model the introduction of linguistic noise due to the presence of a new language. The single grammar equilibrium may become unstable, and when the noise is removed, the population may settle at a completely different equilibrium, thereby mimicking catastrophic changes such as those observed in the transition from Old English to Middle English.

Explorations in low dimensions in Chapter 3 reveal that the language dynamical equation can exhibit regular and chaotic oscillations. These cases of the model directly capture the ability of languages to change spontaneously and unpredictably, while still following a regular pattern. Chaotic behavior displays the same sensitivity to small changes observed in some cases of language contact.

Both of these cases are more realistic than pure replicator dynamics, which can only exhibit fixed stable equilibria for the type of communication game studied here. Thus, the addition of a learning process to game dynamics is crucial for modeling language change.

6.1.2. Modeling change in universal grammar. From there, we extended the model to include genetic variation within universal grammar in the hope of modeling language change on geological timescales. Again due to the unlimited complexity of the language dynamical equation, it was necessary to begin with some low dimensional cases.

Chapter 4 analyzed various competitions between two UGs where one of the UGs admitted only a single grammar. The first set of results assumed that all grammars were unambiguous and individuals with the same grammar communicate perfectly. Under these circumstances, single-grammar UGs are quite stable, apparently due to the fact that their one grammar is learned perfectly. Consequently, they can successfully invade populations using that grammar but with another UG with imperfect learning. However, a population dominated by a sufficiently different grammar is immune to invasion by the single-grammar UG. These results illustrate how a very flexible and less specific UG admitting many languages may potentially be replaced by a more specific UG admitting a subset of those languages. If grammars are allowed to be ambiguous, so that communication between individuals with the same grammar is imperfect, then coexistence of single-grammar UGs and multi-grammar UGs is possible. It is also possible for a more specific grammar to be replaced by a less specific UG if the new grammars are less ambiguous. These results are progress toward understanding how UG balances the benefits of flexibility against the needs for precise communication and accurate learning.

We then examined cases of two UGs with two grammars each. With some symmetry imposed on the parameters, it is possible to solve for the fixed points and (almost) determine their stabilities. One result is that if the two UGs admit sufficiently similar grammars, then they can coexist stably, indicating that the actual human UG may include several very similar genetic variants, all coexisting. Another result is that UGs with sufficiently different grammars are stable against invasion by each other, that is, they exhibit competitive exclusion. This result can be generalized to sufficient conditions on any set of four grammars that imply that the resulting UGs exclude each other. Hence, market share effects are extremely important: Evolution of UG must be incremental, and innovations must maintain a certain level of compatibility with the existing population to be successful. There are also cases where these conditions do not hold, and the model exhibits strange behavior such as accidental stability: One UG may successfully invade the other only if the invaded population is dominated by one grammar and not the other. Thus, some aspects of UG are not consequences of natural selection in any straightforward sense. They were preserved not because they are inherently better for communication in general, but because of some historical accident. They were adaptive in the sense that they facilitated communication in some particular population.

6.2. Consequences for mathematical biology

The interaction of game dynamics and learning raises interesting modeling problems not found in pure replicator dynamics. The primary issue is that the "fitness of a particular universal grammar" is an incomplete concept in that it depends on other information. This is nothing new. Many games, such as hawk and dove [31, 46] are structured so that fitness only makes sense in reference to an opponent or within the context of a population. However, the fitness of a UG depends not only on the frequencies of competing UGs in the population, but also on the frequencies at which grammars are acquired by individuals. UG is one layer removed from the communication game that is the actual source of all payoffs in this scenario.

From another perspective, much of the development of multicellular organisms is self-organizing, driven by a mixture of learning and genetically encoded information. For example, the neurological connection from the eye to the vision centers in the brain is not specified in every detail in the genome. Rather, it develops based on continuity. Adjacent rods and cones in the retina send highly correlated signals to the brain because images of the world are largely continuous. Taking advantage of this fact, the eye–brain connection organizes itself on the basis of which connecting neurons are carrying correlated signals. Likewise, language develops in the brain based on patterns in the linguistic environment. However, the language faculty creates this environment through community, so universal grammar is a self-organizing system like the eye–brain connection, but with an additional source of feedback.

Putting all of this together, there is no way to simplify a game among universal grammars to replicator dynamics as described in Section 7.1 of [31], because the fitness of a particular UG depends on more than the frequencies of other UGs in the population. It also depends on the linguistic environment, which is partially determined by feedback from the UGs in the population, but also depends on history and chance. It is therefore necessary to study games among UGs using the language dynamical equation, as the learning process is a crucial extension to the replicator model.

Mathematically, it is necessary to think of universal grammar as a metastrategy. We suppose that there is an underlying game with some number of strategies available, and this game generates the payoff. Each individual has a metastrategy, that is, a strategy for choosing a strategy for the underlying game. In the communication game, a grammar is a strategy, and a UG, more specifically its learning process, is a metastrategy. To determine whether a particular UG is stable in some sense, it is necessary to use the full power of dynamical systems theory; concepts such as the Nash equilibrium and evolutionary stability are not useful unless they are generalized to account for the additional information required to define the fitness of a UG. The material in Chapters 4 and 5 is progress toward developing methods for analyzing metastrategy games.

6.3. Consequences for linguistics

The use of an evolutionary model to discuss historical language change immediately runs into the problem that according to many linguists, no language in current use is inherently fitter than any other. For example, Lightfoot [45, § 8.2] criticizes a specific argument that head-final Latin developed into head-initial Romance languages such as French, following an evolutionary progression, because head-initial languages are easier to learn and therefore fitter in some sense. Lightfoot claims that changes in primary linguistic data should be the preferred explanation for such changes, and points out that the argument in question fails to explain how Latin came to be head-final in the first place.

Mathematically, the current model avoids this problem in two ways. First, the model defines the fitness of a grammar as a function of the population state and the ability of an individual to communicate within that population. Second, the payoff in the communication game of a grammar playing against itself (a diagonal entry of B) is generally assumed to be the same or nearly the same for each grammar. Thus, the fitness measure of a grammar is not based solely on inherent properties of that grammar, and the one parameter that does measure a property of just that grammar is close to invariant in most of the examples in this dissertation, the exception being ambiguous grammars in Section 4.6 which seem to be necessary to destabilize single-grammar UGs. Elsewhere, no grammar is assumed to be generally fitter in any sense than any other.

Furthermore, the most interesting behaviors of this model, namely oscillations and chaos, are driven by the learning process rather than game dynamics (Chapter 3). In this sense, the model provides a test bed for theories such as those in [44, 45] that changes in primary linguistic data interacting with a particular grammar acquisition algorithm are responsible for language change. With this interpretation, the selection terms of the language dynamical equation are implementing a neutral evolutionary game among equally fit strategies that serves to keep changes introduced by the learning process under control.

The evolution of the language faculty presents other puzzles. The ability to communicate at the level of human language is clearly beneficial, but no other species seems to have anything quite like it. Consequently, linguists have struggled with questions relating to the general architecture of the language faculty. At one time, language was thought of as a monolithic system. Since no part of it could function without the entire rest of the system, an evolutionary explanation seemed out of the question, as a fully functional language faculty would have had to appear all at once. Recently, a more modular theory of language has become popular. Jackendoff [34] proposes a highly modular understanding of language, in which certain parts are useful without the rest, and so may have appeared incrementally.

A related issue is how specialized the language faculty is. Hauser *et al.* [28] discusses three general hypothesis concerning the language faculty in a broad sense (FLB) and in a narrow sense (FLN). FLB contains sensory-motor and conceptual-intentional subsystems, as well as FLN which is the abstract computational system specific to language. One hypothesis is that FLB is built from the same underlying components used for communication in other species. The second is that FLB is uniquely human. The third, most strongly supported by the authors of [28], is that FLB is mostly based on mechanisms shared with other species, but the subset FLN is unique to humans. The third hypothesis allows for the possibility that features useful to language initially evolved because they have other benefits, and were later combined under a specialized computational system only in humans.

The results of Chapters 4 and 5 indicate that the role of natural selection in the evolution of language is not clear. As noted before, there is no simple notion of what makes a UG adaptive, or when one is fitter than another. Market share effects are also extremely important: A linguistic innovation that allows for extremely powerful or efficient communication is useless in an environment where no one else understands it. This can lead to counterintuitive situations where an arguably inferior language faculty is maintained even in the presence of an arguably superior alternative. Furthermore, multiple universal grammars may be able to coexist, and the success of one UG over another may be a consequence of a historical accident. Thus, it is likely that some features of UG are present not because they are adaptive in any simple sense, but because of market share effects, historical accident, physical or mathematical constraints, or other causes.

For example, Lightfoot [45] argues that a constraint within UG on movement is maladaptive and could not have been selected for. The rule forbids certain kinds of questions where the subject of an embedded clause is queried [45, p. 244]:

(6.3.1) I thought that Ray saw Fay.

(6.3.2) Who$_1$ did you think t$_1$ that Ray saw t$_1$?

(6.3.3) *Who$_1$ did you think t$_1$ that t$_1$ saw Fay?

He proceeds to illustrate how a variety of languages have special rules that allow them to by-pass this feature of UG. The rule in English is to use a null complementizer in place of *that:*

(6.3.4) Who$_1$ did you think t$_1$ t$_1$ saw Fay?

Lightfoot concludes that although the rules of government and binding may be adaptive in some sense and were perhaps selected for, the particular consequence that forbids the formation of this class of questions without additional *ad hoc* mechanisms is maladaptive. It is therefore a spandrel, an unintended consequence of some other design decision, and its presence in UG should not be explained in terms of its being adaptive.

Lightfoot [45, § 9.2] also proposes that language evolution is subject to certain mathematical and physical constraints, parallel to constraints that force smaller animals to have faster metabolisms than larger animals. An example of such a constraint might be the need for systems of morphology and phonology when the lexicon exceeds a certain size [63]. So while circumstances might select a large vocabulary over a small one, a combinatorial system is the only way to implement it, and is not selected for in the same sense because there is no alternative. Such constraints should naturally show up in mathematical models, as demonstrated by [63]. As an example from the language dynamical equation from Chapter 2, accurate learning is required for short term stability of grammars. If there is selection for populations where almost everyone uses the same grammar, then accurate learning is inevitable.

6.4. Possible extensions and future work

This research could be extended in any number of directions. Here, I will briefly summarize a few possibilities. The first is a proposal for a detailed discrete simulation of a linguistic population. The second is about connecting the present model to data collected by historical linguists. The remainder are more mathematical, and have to do with studying the language dynamical equation as a dynamical system.

6.4.1. A detailed discrete simulation. A number of questions and criticisms are repeatedly raised concerning linguistics and the assumptions made by this model. Two of these criticisms, regarding the source of learning errors and catastrophic language change, could be addressed by comparing the behavior of the selection/mutation model to a more detailed simulation:

1. Where do these learning errors come from? Most children do a remarkable job of learning the subtleties of their native language, but incremental changes do occur. (For example, compare American English to British English.) The language dynamical equation attributes incremental change to tiny errors that occur during grammar acquisition. A detailed simulation of individuals learning from a population might be able to illustrate and verify the conditions that lead to such errors, such as style and the effect of age groups. Such a simulation could also test proposed acquisition algorithms for phonetic systems, vocabulary, and syntax. (For example, see the RIP/CD algorithm described in [72], zero-degree learning from [44, 45], and parallel tree structures from [34].) It would be computationally intensive compared to the language dynamical equation, which is a system of ordinary differential equations. However, comparison with the simulation can help determine whether the language dynamical equation is at an appropriate level of abstraction for understanding particular aspects of language.

2. How can language contact cause catastrophic changes? A theory mentioned in [45] says that the transformation of Old English into Middle English may have been caused (or accelerated) by the presence of Scandinavian invaders speaking Old Norse. The two languages were sufficiently similar that children were unable to reliably learn Old English due to what might be called linguistic noise. This theory is reasonable in that it cleanly explains how the Old English case system was lost, but paradoxically, children are very good at learning multiple native languages. The resolution of the paradox may be the presence of a phase transition, in which sufficiently different languages are acquired bilingually, but sufficiently similar languages merge into one. Merging happens all the time: At the population level, a language incorporates variation in the form of registers (formal, casual, literary, slang) and dialects (New York City, British, Australian). Children overlook these fine-scale differences and acquire some average form of the language as spoken by the whole population. The phase boundary is in between, when two languages are simultaneously too different to be considered dialects of a single language, yet similar enough that children cannot acquire them bilingually, yielding a catastrophe. Old English and Old Norse may have been near such a boundary. The language dynamical equation and the simulation described above could be modified to incorporate learning from a population, directly model a similar catastrophic change, and perhaps map out the phase boundary.

6.4.2. Connecting the language dynamical equation to data. These first stages in the analysis of the language dynamical equation have focused on phenomena, such as coherent and incoherent equilibria, oscillations, and chaos. This model would be more useful to linguists if it could make testable predictions concerning theories of syntax and language acquisition.

As an example of what could be done, consider the cue-based acquisition model proposed by Lightfoot [44, 45]. The learning process in the language dynamical equation could be modified as follows to model this proposal. Suppose that each grammar produces sentences of particular kinds at certain rates, and that children are sensitive to the overall rate at which those sentences appear in the environment. The modified learning process would set parameters based on which kinds of sentences occur at a rate above some threshold. Analysis of the resulting dynamical system could provide an indication of what perturbations of the linguistic environment are sufficiently large to trigger a population-wide shift from one grammar to another. The rate at which that change occurs could also be tracked and compared with actual data, such as [40]. If the dynamical system agrees with the data, then the model of acquisition is supported. If the dynamical system cannot be made to agree with the data, then the model of acquisition may need to be revised.

In short, appropriate modifications to the language dynamical equation can convert it to a testing framework for particular models of language acquisition.

6.4.3. Dynamical systems questions.

6.4.3.1. *More about chaos.* As described in Chapter 3, the language dynamical equation exhibits period doubling and chaotic behavior. The parameter settings in question create orbits that spiral inward, then escape in a new direction, and finally return to the spiral. There are any number of further questions that could be asked about this example. To begin with, it is likely that some theorems by Šilnkov concerning saddle-foci apply to the example [26, 75]. Since the escape-and-return mechanism can be constructed in any number of dimensions, it may be necessary to generalize some of those results (which have been proved in three or four dimensions) before they can be applied to the example. Furthermore, the Poincaré map shown in Section 3.3 appears to exhibit a three-dimensional generalization of Smale's horseshoe [14], and some effort could be put into developing the details.

It appears possible to use multiple escape-and-return mechanisms to construct arbitrarily complex chaotic orbits. The general idea is to link several spirals together, such that orbits have at least two ways to escape out of any given spiral. The result would be chaotic orbits that switch from one spiral to another unpredictably, but with some control over which spirals can be reached by which others. This construction may be a useful source of examples of chaotic behavior. A few preliminary experiments have not been successful at creating such orbits, but more trials should be done.

6.4.3.2. *Structural stability of replicator dynamics.* It seems quite likely that most cases of replicator dynamics with a diagonally dominant payoff matrix are structurally stable. If such a theorem were to be proved, it would imply that small perturbations of the vector field would result in essentially the same phase portrait: stable fixed points near the corners of the simplex. For sufficiently small learning errors, the language dynamical equation could be viewed as a perturbation of the replicator dynamics. It would follow that for nearly perfect learning, populations near the corners always converge to a nearby stable fixed point.

6.4.3.3. *Competition between learning processes.* The results of Section 5.4 apply to contests between UGs that satisfy certain constraints. If the languages they admit are sufficiently different, then they are stable against invasion by one another. An interesting follow-up question is to ask what happens when the two UGs admit exactly the same grammars, but have different learning algorithms. The resulting dynamical system appears to be genuinely difficult to analyze, as the z null-cline argument is of no use, and alternative methods must be found.

6.5. Last words

This dissertation has illustrated that the unexpected combination of mathematics and linguistics has a lot to offer to both fields. It is my hope that the field of mathematical linguistics will continue to grow, and produce an ever richer understanding of communication, community, learning, and history.

Bibliography

[1] L. V. Ahlfors. *Complex Analysis*. McGraw-Hill, third edition, 1979.

[2] J. Aitchinson. *Words in the Mind: An Introduction to the Mental Lexicon*. Basil Blackwell, Oxford, 1987.

[3] A. A. Andronov, E. A. Leontovich, I. I. Gordon, and A. G. Maier. *Theory of Bifurcations of Dynamic Systems on a Plane*. Keter Press, Jerusalem, 1971.

[4] D. Angluin. Learning regular sets from queries and counterexamples. *Information and Computation*, 75:87–106, 1987.

[5] D. Angluin and M. Kharitonov. When won't membership queries help? *Journal of Computer and System Sciences*, 50(336–355), 1995.

[6] P. Ashwin and J. W. Swift. The dynamics of n weakly coupled identical oscillators. *Journal of Nonlinear Science*, 2:69–108, 1992.

[7] D. Bickerton. *Roots of Language*. Karoma Publishers, Inc., Ann Arbor, 1981.

[8] D. Bickerton. *Language and Species*. University of Chicago Press, Chicago, 1990.

[9] A. Cangelosi and D. Parisi. The emergence of a "language" in an evolving population of neural networks. *Connection Science*, 10(2):83–97, 1998.

[10] A. Cangelosi and D. Parisi, editors. *Simulating the Evolution of Language*. Springer-Verlag, 2001.

[11] N. Chomsky. *Language and Mind*. Harcourt Brace Jovanovich, New York, 1972.

[12] N. Chomsky. *Language and Problems of Knowledge*. MIT Press, 1988.

[13] T. Crowley. *An Introduction to Historical Linguistics*. Oxford Univeristy Press, third edition, 1998.

[14] R. L. Devaney. *An Introduction to Chaotic Dynamical Systems*. Addison-Wesley, second edition, 1989.

[15] M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, 1971.

[16] 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.

[17] M. Eigen and P. Schuster. *The Hypercycle: A Principle of Natural Self-Organisation*. Springer-Verlag, Berlin, 1979.

[18] W. Enard, M. Przeworski, S. E. Fisher, C. S. L. Lai, V. Wiebe, T. Kitano, A. P. Monaco, and S. Pääbo. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, 418(6900):869–872, 2002.

[19] R. Ferrer i Cancho and R. V. Solé. Two regimes in the frequency of words and the origin of complex lexicons: Zipf's law revisited. *Journal of Quantitative Linguistics*, 8:165–173, 2001.

[20] 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.

[21] S. A. Frank. *Foundations of Social Evolution*. Princeton University Press, Princeton, NJ, 1998.

[22] A. A. Ghazanfar and M. D. Hauser. The neuroethology of primate vocal communication: substrates for the evolution of speech. *Trends in Cognitive Sciences*, 3(18):377–384, 1999.

[23] E. Gibson and K. Wexler. Triggers. *Linguistic Inquiry*, 25:407–454, 1994.

[24] E. M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

[25] N. Grassly, A. von Haesler, and D. C. Krakauer. Error, population structure and the origin of diverse sign systems. *Journal of Theoretical Biology*, 206:369–378, 2000.

[26] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer-Verlag, 1990.

[27] M. D. Hauser. *The Evolution of Communication*. Harvard University Press, Cambridge, MA, 1996.

[28] M. D. Hauser, N. Chomsky, and W. T. Fitch. The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598):1569–1579, 2002.

[29] M. D. Hauser, E. L. Newport, and R. N. Aslin. Segmentation of the speech stream in a nonhuman primate: statistical learning in cotton-top tamarins. *Cognition*, 78(3):B53–B64, 2001.

[30] M. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, 1997.

[31] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

[32] J. R. Hurford, M. Studdert-Kennedy, and C. Knight, editors. *Approaches to the Evolution of Language*. Cambridge University Press, 1998.

[33] R. Jackendoff. Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, 3:272–279, 1999.

[34] R. Jackendoff. *Foundations of Language*. Oxford University Press, Oxford, 2002.

[35] S. Kirby. Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2):102–110, 2001.

[36] N. L. Komarova, P. Niyogi, and M. A. Nowak. The evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology*, 209(1):43–59, 2001.

[37] N. L. Komarova and M. A. Nowak. The evolutionary dynamics of the lexical matrix. *Bulletin of Mathematical Biology*, 63(3):451–485, 2001.

[38] N. L. Komarova and M. A. Nowak. Natural selection of the critical period for language acquisition. *Proceedings of the Royal Society of London, Series B*, 268:1189–1196, 2001.

[39] D. C. Krakauer. Kin imitation for a private sign system. *Journal of Theoretical Biology*, 213:145–157, 2001.

[40] A. Kroch. Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1:199–244, 1989.

[41] M. Lachmann, S. Szamado, and C. T. Bergstrom. Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences, USA*, 98:13189–94, 2001.

[42] C. S. L. Lai, S. E. Fisher, J. A. Hurst, F. Vargha-Khadem, and A. P. Monaco. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, 413(6855):519–523, 2001.

[43] P. Lieberman. *The Biology and Evolution of Language*. Harvard University Press, Cambridge, MA, 1984.

[44] D. Lightfoot. *How to Set Parameters: Arguments from Language Change*. MIT Press, Cambridge, MA, 1991.

[45] D. Lightfoot. *The Development of Language: Acquisition, Changes and Evolution*. Blackwell Publishers, 1999.

[46] R. M. May. *Stability and Complexity in Model Ecosystems*. Princeton University Press, Princeton, NJ, 2001.

[47] J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, 1982.

[48] W. G. Mitchener. Bifurcation analysis of the fully symmetric language dynamical equation. *Journal of Mathematical Biology*, 46:265–285, 2003.

[49] W. G. Mitchener and M. A. Nowak. Competitive exclusion and coexistence of universal grammars. *Bulletin of Mathematical Biology*, 65(1):67–93, 2003.

[50] P. Niyogi. *The Informational Complexity of Learning*. Kluwer Academic Publishers, Boston, 1998.

[51] P. Niyogi and R. C. Berwick. A language learning model for finite parameter spaces. *Cognition*, 61:161–193, 1996.

[52] P. Niyogi and R. C. Berwick. Evolutionary consequences of language learning. *Linguistics and Philosophy*, 20:697–719, 1997.

[53] M. Nowak and K. Sigmund. Chaos and the evolution of cooperation. *Proceedings of the National Academy of Sciences, USA*, 90:5091–5094, 1993.

[54] M. A. Nowak and N. L. Komarova. Towards an evolutionary theory of language. *Trends in Cognitive Sciences*, 2001.

[55] M. A. Nowak, N. L. Komarova, and P. Niyogi. Evolution of universal grammar. *Science*, 291(5501):114–118, 2001.

[56] M. A. Nowak, N. L. Komarova, and P. Niyogi. Computational and evolutionary aspects of language. *Nature*, 417(6889):611–617, 2002.

[57] M. A. Nowak and D. C. Krakauer. The evolution of language. *Proceedings of the National Academy of Sciences, USA*, 96:8028–8033, 1999.

[58] M. A. Nowak, J. Plotkin, and V. A. A. Jansen. Evolution of syntactic communication. *Nature*, 404(6777):495–498, 2000.

[59] M. A. Nowak, J. Plotkin, and D. C. Krakauer. The evolutionary language game. *Journal of Theoretical Biology*, 200:147–162, 1999.

[60] K. M. Page and M. A. Nowak. Unifying evolutionary dynamics. *Journal of Theoretical Biology*, 219:93–98, 2002.

[61] S. Pinker. *The Language Instinct*. W. Morrow and Company, New York, 1990.

[62] S. Pinker and A. Bloom. Natural language and natural selection. *Behavioral and Brain Sciences*, 13:707–784, 1990.

[63] J. Plotkin and M. A. Nowak. Language evolution and information theory. *Journal of Theoretical Biology*, 205:147–159, 2000.

[64] G. R. Price. Extension of covariance mathematics. *Annals of Human Genetics*, 35:485–490, 1972.

[65] F. Ramus, M. D. Hauser, C. Miller, D. Morris, and J. Mehler. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288(5464):349–351, 2000.

[66] W. Schnabel, P. F. Stadler, C. Forst, and P. Schuster. Full characterization of a strange attractor. *Physica D*, 48(1):65–90, 1991.

[67] P. Schuster. Structure and dynamics of replication-mutation systems. *Physica Scripta*, 35(3):402–416, 1987.

[68] A. Senghas and M. Coppola. Children creating language: How Nicaraguan sign language acquired a spatial grammar. *Psychological Science*, 12(4):323–328, 2001.

[69] P. F. Stadler and P. Schuster. Mutation in autocatalytic reaction networks. *Journal of Mathematical Biology*, 30(6):597–632, 1992.

[70] S. H. Strogatz. *Nonlinear Dynamics and Chaos*. Perseus Books, Reading, MA, 1994.

[71] M. Studdert-Kennedy. Evolutionary implications of the particulate principle: Imitation and the dissociation of phonetic form from semantic function. In C. Knight, J. R.

Hurford, and M. Studdert-Kennedy, editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge University Press, Cambridge, 2000.

[72] B. Tesar and P. Smolensky. *Learnability in Optimality Theory*. MIT Press, 2000.

[73] P. E. Trapa and M. A. Nowak. Nash equilibria for an evolutionary language game. *Journal of Mathematical Biology*, 41:172–1888, 2000.

[74] R. L. Trask. *Historical Linguistics*. Arnold, London, 1996.

[75] C. Tresser. About some theorems by L. P. Šil'nikov. *Annales de l'Institute Henri Poincaré – Physique théorique*, 40(4):442–461, 1984.

[76] J. Uriagereka. *Rhyme and Reason: An Introduction to Minimalist Syntax*. MIT Press, Cambridge, MA, 1998.

[77] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27:436–445, 1984.

[78] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.

[79] T. Warnow. Mathematical approaches to comparative linguistics. *Proceedings of the National Academy of Sciences, USA*, 94:6585–6590, 1997.

[80] C. Watkins, editor. *The American Heritage Dictionary of Indo-European Roots*. Houghton Mifflin, second edition, 2000.