Conditions for Global Asymptotic Dominance in Multidimensional Grammar Competition

Henri Kauhanen
Department of Linguistics, University of Konstanz
henri.kauhanen@uni-konstanz.de

Abstract

In variational learning of two grammars, the grammar with the greater advantage is globally asymptotically dominant, meaning that this grammar wins out over inter-generational time, leading to the extinction of its competitor. Here, we prove necessary and sufficient conditions for global asymptotic dominance in competition between an arbitrary number $n \geq 2$ of grammars, working in the deterministic limit of iterated inter-generational linear reward–penalty learning.

1 Introduction

Grammar competition (Kroch, 1989), particularly when combined with variational learning[1] (Yang, 2002), is a much-employed framework for explaining language change (see e.g. Yang, 2000; Heycock and Wallenberg, 2013; Simonenko et al., 2019). In its usual application, the procedure is to define a simple model of language acquisition, and then to bootstrap a model of population-level linguistic change from it. This is typically done by setting up a discrete sequence of non-overlapping generations of learners, assuming that the learning environment for learners in each generation is constituted by the average linguistic production of learners of the immediately preceding generation, abstracting away from stochastic fluctuations (cf. Andersen, 1973).

The dynamics of such a system are fully understood in the case of two competing grammars; in this special case, the population-level equation is a replicator dynamic with a flat fitness landscape. As a consequence, these systems are capable of two types of behaviour only: either every population state is a stable (albeit not asymptotically stable) equilibrium, or a single asymptotically stable equilibrium exists, attracting all non-equilibrium initial states. Typically, the latter situation is the linguistically interesting one, implying that, over repeated inter-generational interactions, one grammar ousts the other. Which grammar wins depends on the competing grammars' relative advantages; these can be estimated from corpora, enabling one to empirically test the predictions made by the mathematical model.

The linear reward–penalty learning scheme (Bush and Mosteller, 1955) underlying the variational learner has a natural extension to n actions—in our case, to competition between an arbitrary number n of grammars. This extension has rarely been employed, however, and its mathematics also remain—apart from results concerning systems embodying certain kinds of strong symmetries (Kauhanen, 2019)—unexplored.

Here, we study the general n-grammar model in detail. We demonstrate that the population-level equation is again a replicator dynamic, although generically (whenever $n > 2$) fitnesses are frequency-dependent and nonlinear. Despite this, certain crucial aspects of the dynamics remain characterizable. In particular, we prove necessary and sufficient conditions for the *global asymptotic dominance* of a single grammar under n-way competition, meaning that the population state corresponding to total use of this grammar is an attractor for any non-equilibrium initial state in which the grammar has some non-zero abundance. In other words, we provide an answer to the question: once a particular grammatical change has been actuated (cf. Weinreich et al., 1968), under what conditions will it proceed to completion? These conditions only reference pairwise advantages between grammars and as such they generalize the well-known "fundamental theorem of language change" for two grammars (Yang, 2000). Our results turn on the notion of the *resilience* of a grammar, defined as the reciprocal of the advantage that its competitor(s) hold(s) over it.

[1]Not to be confused with the collection of methods known as variational inference in Bayesian learning theory (MacKay, 2003).

The necessary and sufficient conditions for global asymptotic dominance make available further results about the dynamics of grammar competition between more than two grammars. As an illustration, we demonstrate how the sufficient condition can alternatively be characterized using the notion of grammatical *flux*, a measure of the tendency for a grammar to lose abundance to its competitors in a speech community.

2 Definitions

In the general case, we assume that the language learner has access to n grammars $G_1, \ldots, G_n$ and attaches a *weight* W_i to each grammar G_i (cf. Yang, 2002). The weights are assumed to form a categorical probability distribution over the grammars; in other words, they must satisfy the following conditions:[2]

1. $W_i \geq 0$ for all i
2. $\sum_i W_i = 1$

Each weight is a random variable; together they form the random probability vector $\mathbf{W} = [W_1 \ldots W_n]^T = (W_1, \ldots, W_n)$. We note that $\mathbf{W}$ is an element of Δ^{n-1}, the simplex

$$\Delta^{n-1} = \left\{\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}_+^n : \textstyle\sum_i x_i = 1\right\}.$$

Whenever no confusion can arise, Δ^{n-1} will be written as Δ for short. This set is partitioned into its *boundary*

$$\partial\Delta = \{\mathbf{x} \in \Delta : x_i = 0 \text{ for some } i\}$$

and *interior*

$$\Delta^\circ = \{\mathbf{x} \in \Delta : x_i > 0 \text{ for all } i\}.$$

The boundary $\partial\Delta$ contains, in particular, the n *vertices*, which are the standard basis vectors of $\mathbb{R}^n$; we denote these by $\mathbf{e}_1 = (1, 0, \ldots, 0)$ through to $\mathbf{e}_n = (0, \ldots, 0, 1)$.

Additionally, for every non-empty subset of grammar indices $H \subseteq \{1, \ldots, n\}$, we define Δ_H as the set of points $\mathbf{x} \in \Delta$ in which G_i has zero weight for each $i \notin H$, in other words

$$\Delta_H = \{\mathbf{x} \in \Delta : x_i = 0 \text{ for all } i \notin H\}.$$

Equivalently, $\Delta_H = \Delta \cap \operatorname{span}\{\mathbf{e}_i : i \in H\}$. Intuitively, Δ_H is the set of possible weight vectors in which grammars outside the index set H never occur; it is itself an $(|H| - 1)$-dimensional simplex.

Each grammar is assumed to accept a language, understood as a set of grammatical expressions out of all expressions it is possible to form using a common alphabet Σ. In other words, to each G_i we attach a set L_i satisfying $L_i \subseteq \Sigma^*$ where Σ^* is the Kleene star (the infinite set of all finite sequences of symbols from Σ). If $s \in L_i$ (i.e. if G_i accepts the string s), we also express this as $G_i \vdash s$. The complement of L_i with respect to Σ^* will be written $\complement L_i$; for all $s \in \complement L_i$, we have $G_i \nvdash s$.

To each G_i, we attach a probability measure μ_i on Σ^* with support on L_i, referred to as the grammar's associated *production measure*. For any $K \subseteq \Sigma^*$, $\mu_i(K)$ gives the probability of G_i producing an expression belonging to K. Whereas acceptance ($\vdash$) is a purely formal, grammatical property, production (μ_i) may depend on both grammatical and extra-grammatical factors, the latter including factors such as discourse constraints and performance limitations.[3]

Expanding on Yang (2000), and following Kauhanen (2019), we adopt the following definition.

Definition 1. The quantity

$$a_{ij} = \mu_j(\complement L_i)$$

is called the *(pairwise) advantage* of grammar G_j over grammar G_i.

In words, a_{ij} is the probability of G_j producing a string that G_i does not accept. It is useful to collect pairwise advantages in an $n \times n$ matrix $A = [a_{ij}]$; we refer to this as an *advantage matrix*. Note that such matrices have zero diagonals: $a_{ii} = 0$ since, by assumption, each production measure μ_i has no support outside L_i.

A learner operates in a learning environment, which we take to be a probability measure on Σ^* according to which sequences of expressions are presented to the learner.

Definition 2. A *learning environment* is any probability measure μ on Σ^*.

[2] Throughout this paper, missing bounds such as on a summation ($\sum_i$) are taken to imply that the index ranges from 1 to n, i.e. over all grammars. Further conditions on indices will be explicitly indicated whenever necessary.

[3] Construction of production measures is non-trivial. Under traditional conceptions of grammar, the language L_i generated by a grammar G_i is typically a (countably) infinite set, containing, among other things, expressions whose production probabilities are practically indistinguishable from zero (e.g. centre-embeddings to arbitrary depth). In practice, it normally suffices to assume that μ_i has finite support. In any case, these complexities need not concern us here, as long as *some* well-posed probability measure μ_i exists for each G_i.

Definition 3. The quantity

$$c_i^\mu = \mu(\complement L_i)$$

is referred to as the *penalty probability* of grammar G_i in the environment μ.

Whenever no confusion can arise, we drop the measure μ from the notation and denote the penalty probability of G_i simply as c_i.

Note that, on this definition, a learning environment is necessarily stationary, as it is associated with a single, unchanging measure on Σ^*. In practice, this means we assume that each learner (in a given generation of learners) is exposed to a constant set of penalty probabilities for the competing grammars over the duration of learning.

An important distinction exists between learning environments that punish each grammar at least some of the time, and environments which never punish some grammar(s). In order to have terminology for these, we define:

Definition 4. A learning environment is *omnipunitive* if $c_i > 0$ for all i. If this is not the case (i.e. if $c_i = 0$ for at least one i), the learning environment is *parapunitive*.

Now assume the learner receives an infinite random sequence of expressions $s_1, s_2, s_3, \ldots \in \Sigma^*$ according to such a μ. Upon reception of s_m, the learner samples a grammar to employ according to the current value of $\mathbf{W}$ (i.e. G_i is chosen with probability W_i). Suppose this grammar is G_k. The learner then checks whether $G_k \vdash s_m$ and, based on this result, applies one or another operator to $\mathbf{W}$ to transform it to its new value, $\mathbf{W}'$.[4] This amounts to assuming the existence of n *reward operators* $u_i^+ : \Delta \longrightarrow \Delta$ and n *punishment operators* $u_i^- : \Delta \longrightarrow \Delta$ and setting

$$\mathbf{W}' = \begin{cases} u_k^+(\mathbf{W}) & \text{if } G_k \vdash s_m \\ u_k^-(\mathbf{W}) & \text{if } G_k \nvdash s_m \end{cases} \tag{1}$$

The learner then receives the next expression in the sequence (s_{m+1}), and the cycle continues.

The central learning-theoretic challenge concerns whether $\mathbf{W}$ converges in some relevant sense as learning continues, based on particular choices for the reward (u_i^+) and punishment (u_i^-) operators, subject to a given learning environment μ. Convergence results exist for a number of choices of these operators (see Bush and Mosteller, 1955; Norman, 1972; Narendra and Thathachar, 1989); in what follows, we will make use of these results in order to study a deterministic approximation, or *mean dynamic* (Sandholm, 2010), of the population-level evolution of a sequence of generations of such learners.

To obtain the population-level mean dynamic, we first need to obtain a mean dynamic that expresses how the expected value of the learner's weight vector, $\overline{\mathbf{W}} = \mathrm{E}[\mathbf{W}]$, evolves. In general, from (1) we have that

$$\overline{W}_i' = \sum_j W_j \big(c_j u_j^-(\mathbf{W})_i + (1 - c_j) u_j^+(\mathbf{W})_i\big).$$

Taking expectations on both sides, this becomes

$$\overline{W}_i' = \sum_j \overline{W}_j \big(c_j u_j^-(\overline{\mathbf{W}})_i + (1 - c_j) u_j^+(\overline{\mathbf{W}})_i\big). \tag{2}$$

In the following sections, the operators $\{u_i^+, u_i^-\}$ will be given particular forms and the resulting mean dynamic studied in detail.

Passage to the population-level mean dynamic then follows via one (or both) of two routes. One can either assume that learning is so slow (but continued for long enough) that stochastic fluctuations inherent in the learning process affect the learner's eventual weight vector $\mathbf{W}$ only to an insignificant extent: the actual value of $\mathbf{W}$ stays close to its expected value, $\overline{\mathbf{W}}$ (see Norman, 1972; Narendra and Thathachar, 1989). Alternatively (or additionally), one can assume that the learners in each generation randomly sample input from an infinite population of independent, identically distributed learners in the previous generation; adherence to the population-level mean dynamic then follows by the Law of Large Numbers (cf. Niyogi, 2006). Taking either of these limits (learning rate to zero, or population size to infinity) recovers the same mean dynamic that relates the competing grammars' abundances from one generation to the next.

3 Linear Reward–Penalty Learning of Two Grammars

In the classical two-action learning problem (i.e. when $n = 2$), it has been customary ever since Bush and Mosteller (1955) to employ the linear map described by the matrix

$$U_1^+ = \begin{bmatrix} 1 & \gamma \\ 0 & 1-\gamma \end{bmatrix}$$

[4] Throughout this paper, we use the notation x' for the successor of x.

for the reward operator u_1^+, meaning that

$$u_1^+(\mathbf{W}) = U_1^+\mathbf{W} = \begin{bmatrix} W_1 + \gamma W_2 \\ W_2 - \gamma W_2 \end{bmatrix}.$$

To punish G_1, the matrix

$$U_1^- = \begin{bmatrix} 1-\gamma & 0 \\ \gamma & 1 \end{bmatrix}$$

is used, so that

$$u_1^-(\mathbf{W}) = \begin{bmatrix} 1-\gamma & 0 \\ \gamma & 1 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = \begin{bmatrix} W_1 - \gamma W_1 \\ W_2 + \gamma W_1 \end{bmatrix}.$$

The corresponding matrices for rewarding and punishing G_2 are obtained as row permutations of U_1^+ and U_1^-:

$$U_2^+ = \begin{bmatrix} 0 & 1-\gamma \\ 1 & \gamma \end{bmatrix} \quad \text{and} \quad U_2^- = \begin{bmatrix} \gamma & 1 \\ 1-\gamma & 0 \end{bmatrix}.$$

In each of these formulae, $0 < \gamma < 1$ is a parameter that sets the *learning rate*; note we assume that the same learning rate applies to each operator.

Suppose that at least one of the penalty probabilities c_1 and c_2 is strictly positive. Then it is well known (see Bush and Mosteller, 1955; Narendra and Thathachar, 1989) that the expected value $\overline{\mathbf{W}}$ tends to

$$\overline{\mathbf{W}}^* = \left(\frac{c_2}{c_1 + c_2}, \frac{c_1}{c_1 + c_2} \right) \tag{3}$$

with increasing learning iteration. In other words, $\overline{\mathbf{W}}^*$ is the (unique) fixed point of the learner's mean dynamic in this case.

Suppose we have a generation of learners, all exposed to the same learning environment, and sample a learner at random from this population. What is the probability that this learner employs G_1? Since all learners are assumed to operate in the same learning environment, this probability is identical to the probability of a single learner employing G_1 which, in turn, is just the expected value $\overline{\mathbf{W}}^*$. To avoid a proliferation of complex notation, and also to consistently distinguish between the individual and population levels, in what follows we will write x_i for the probability of encountering an individual employing grammar G_i, and call this the *abundance* of G_i. The abundances can be collected in a vector, $\mathbf{x} = (x_1, \ldots, x_n)$; of course, $\mathbf{x} \in \Delta$.

At the population level, the probability measure μ which defines the learning environment can be decomposed into the individual production measures of the competing grammars. Specifically, for any $K \subseteq \Sigma^*$,

$$\mu(K) = x_1\mu_1(K) + x_2\mu_2(K).$$

In other words, the probability of encountering a string $s \in K$ equals the probability of meeting a speaker employing G_1 and this speaker producing $s \in K$, plus the probability of meeting a speaker employing G_2 and this speaker producing $s \in K$. By the definition of penalty probability, we then have

$$\begin{cases} c_1 = \mu(\complement L_1) = x_2\mu_2(\complement L_1) = a_{12}x_2, \\ c_2 = \mu(\complement L_2) = x_1\mu_1(\complement L_2) = a_{21}x_1. \end{cases}$$

By (3), a learner in such an environment will tend to converge to

$$(\overline{W}_1^*, \overline{W}_2^*) = \left(\frac{a_{21}x_1}{a_{12}x_2 + a_{21}x_1}, \frac{a_{12}x_2}{a_{12}x_2 + a_{21}x_1} \right).$$

Assuming a discrete sequence of generations of this kind, we thus have the following deterministic difference equation for the abundances of the two grammars in the population:

$$\begin{cases} x_1' = \dfrac{a_{21}x_1}{a_{12}x_2 + a_{21}x_1} \\ x_2' = \dfrac{a_{12}x_2}{a_{12}x_2 + a_{21}x_1} \end{cases}$$

Since $x_1 + x_2 = 1$, it of course suffices to track the evolution of $x = x_1$ only, whereby we have the simple expression

$$x' = \frac{a_{21}}{\alpha(x)} x, \tag{4}$$

where

$$\alpha(x) = a_{12}(1 - x) + a_{21}x$$

denotes average advantage.

Equation (4) is a discrete-time constant-fitness Maynard Smith replicator dynamic (cf. Sandholm, 2010). As such, its behaviour is extremely simple. Outside of the case of a symmetric advantage matrix ($a_{12} = a_{21}$), in which case every $x \in [0, 1]$ is a fixed point, the system has a single asymptotically stable equilibrium that attracts from any initial state $0 < x < 1$. If $a_{21} > a_{12}$, this equilibrium is $x = 1$; if $a_{21} < a_{12}$, it is $x = 0$.[5] This is summarized in:

Theorem 1 (Yang, 2000)**.** *When $n = 2$, the grammar with the greater advantage wins.*

[5] We do not concern ourselves with the pathological case $a_{12} = a_{21} = 0$ in which neither grammar ever produces output that its competitor cannot parse, i.e. in which the grammars are extensionally equivalent.

4 Linear Reward–Penalty Learning of n Grammars

We now turn to the general case of n competing grammars, where Bush and Mosteller's (1955) linear scheme takes the following form. Suppose G_k is the grammar selected by the learner for input sequence s_m. Then, the grammar weights are updated as follows:[6]

$$\begin{aligned} &\text{if } G_k \vdash s_m\text{:} && \begin{cases} W'_k = (1-\gamma)W_k + \gamma \\ W'_i = (1-\gamma)W_i & (i \neq k) \end{cases} \\ &\text{if } G_k \nvdash s_m\text{:} && \begin{cases} W'_k = (1-\gamma)W_k \\ W'_i = (1-\gamma)W_i + \frac{\gamma}{n-1} & (i \neq k) \end{cases} \end{aligned} \tag{5}$$

To characterize a learner's behaviour under this learning algorithm, we focus on the expected weight vector $\overline{\mathbf{W}} = \mathrm{E}[\mathbf{W}]$. With the operators (5), it can be checked that the learner's mean dynamic (2) simplifies to

$$\overline{W}'_i = (1-\gamma c_i)\overline{W}_i + \sum_{j\neq i} \frac{\gamma}{n-1} c_j \overline{W}_j.$$

In other words, we may write

$$\overline{\mathbf{W}}' = B\overline{\mathbf{W}}, \tag{6}$$

where $B = [b_{ij}]$ is the $n \times n$ square matrix with

$$b_{ij} = \begin{cases} 1-\gamma c_i & \text{if } i = j, \\ \frac{\gamma}{n-1} c_j & \text{if } i \neq j. \end{cases}$$

Generically, $\overline{\mathbf{W}}$ evolves to a limit whose value can be obtained by solving the system of equations $B\overline{\mathbf{W}} = \overline{\mathbf{W}}$. In an omnipunitive environment (i.e. $c_i > 0$ for all i), this system of equations has the unique solution $\overline{\mathbf{W}} = (c_1^{-1}/C, \ldots, c_n^{-1}/C)$ with $C = \sum_i c_i^{-1}$ (Bush and Mosteller, 1955; Narendra and Thathachar, 1989).[7] The following result generalizes this to also cover parapunitive environments.

Theorem 2. *If the learning environment is omnipunitive, then*

$$\overline{\mathbf{W}}^* = \left(\frac{c_1^{-1}}{\sum_i c_i^{-1}}, \ldots, \frac{c_n^{-1}}{\sum_i c_i^{-1}} \right)$$

is the only fixed point of the learner's mean dynamic (6).

If it is parapunitive so that $c_i = 0$ for $i \in H_0$, then all $\overline{\mathbf{W}} \in \Delta_{H_0}$ are fixed points.

Proof. First suppose the environment is omnipunitive. Then $c_i > 0$ for all i, and so the matrix B is both positive and column-stochastic. By the Perron–Frobenius Theorem, there is then a unique stationary $\overline{\mathbf{W}}$ that satisfies the equilibrium condition $\overline{\mathbf{W}} = B\overline{\mathbf{W}}$. It is quick to check that the $\overline{\mathbf{W}}^*$ stated in the theorem satisfies it.

Then suppose the environment is parapunitive. We begin by showing that a $\overline{\mathbf{W}}$ that satisfies the conditions stated in the theorem is a fixed point of the mean dynamic. We have

$$\overline{W}'_i - \overline{W}_i = -\gamma c_i \overline{W}_i + \sum_{\substack{j\neq i \\ j\in H_0}} \frac{\gamma c_j}{n-1}\overline{W}_j + \sum_{\substack{j\neq i \\ j\notin H_0}} \frac{\gamma c_j}{n-1}\overline{W}_j.$$

On the right hand side, the second term is zero since all the c_j are zero (since $j \in H_0$), and the third term is zero because all the $\overline{W}_j$ are zero (since $j \notin H_0$). Hence

$$\overline{W}'_i - \overline{W}_i = -\gamma c_i \overline{W}_i.$$

Now, if $i \in H_0$, then $c_i = 0$ and so $\overline{W}'_i - \overline{W}_i = 0$, and we are done. On the other hand, if $i \notin H_0$, then $\overline{W}_i = 0$ and again $\overline{W}'_i - \overline{W}_i = 0$.

To see that the assumed form of $\overline{\mathbf{W}}$ is necessary, suppose $i \in H_0$. Then $c_i = 0$, and so

$$\overline{W}'_i - \overline{W}_i = \sum_{\substack{j\neq i \\ j\notin H_0}} \frac{\gamma}{n-1} c_j \overline{W}_j.$$

In the summation, we have $c_j > 0$ since $j \notin H_0$. Hence, if we had $\overline{W}_j > 0$ for even one j, we would obtain that $\overline{W}'_i - \overline{W}_i \neq 0$. Hence, we must have $\overline{W}_j = 0$ for all $j \notin H_0$, and so $\overline{\mathbf{W}} \in \Delta_{H_0}$. □

[6] Just as in the two-grammar case, this definition gives rise to a set of linear operators. To see this, notice that the first equation, for instance, can be rewritten as

$$W'_k = W_k + \gamma(1 - W_k) = W_k + \gamma \textstyle\sum_{i\neq k} W_i$$

so that W'_k is found to be a linear combination of $W_1, \ldots, W_n$. In fact, if I denotes the $n \times n$ identity matrix (i.e. I has 1 in each diagonal cell and 0 in each non-diagonal cell), J denotes the $n \times n$ matrix of ones (i.e. J has 1 in each cell), and E_{kk} denotes the matrix unit (i.e. E_{kk} has a 1 in cell (k,k) and is 0 elsewhere), then the learning operators can be written concisely as the following matrices:

$$\begin{cases} U_k^+ = (1-\gamma)I + \gamma E_{kk} J \\ U_k^- = (1-\gamma)I + \dfrac{\gamma}{n-1}(J - E_{kk}J) \end{cases}$$

[7] Note that (3) is a special case: we have

$$\overline{W}_1 = \frac{c_1^{-1}}{c_1^{-1} + c_2^{-1}} = \frac{c_1 c_2 c_1^{-1}}{c_1 c_2 (c_1^{-1} + c_2^{-1})} = \frac{c_2}{c_2 + c_1}$$

and similarly for $\overline{W}_2$.

Corollary 1. *If the learning environment is parapunitive with $c_k = 0$ for a unique k, then $\overline{\mathbf{W}}$ tends to the vertex $\mathbf{e}_k$.*

Proof. This follows because $\Delta_{\{k\}}$ is a singleton: $\Delta_{\{k\}} = \Delta \cap \operatorname{span}\{\mathbf{e}_k\} = \{\mathbf{e}_k\}$. □

See Figure 1 for illustration.

Moving to the population level, we note that the learning environment μ again decomposes: for any $K \subseteq \Sigma^*$, we have

$$\mu(K) = \sum_j x_j \mu_j(K).$$

Hence

$$c_i = \mu(\complement L_i) = \sum_j x_j \mu_j(\complement L_i) = \sum_j a_{ij} x_j,$$

and so the penalty probabilities are linear combinations of the grammar abundances. Assuming as before a discrete sequence of infinite generations, we have

$$x_i' = \frac{c_i^{-1}}{\sum_j c_j^{-1}}$$

if $c_i > 0$ for all i, i.e. if the learning environment is omnipunitive at the population state $\mathbf{x}$. As will be explained in the following section, this is usually the case.

In particular, if omnipunitivity holds, then in the interior Δ° (i.e. when $x_i > 0$ for all i), the above equation may be expressed in more familiar terms: multiplying the numerator by $x_i^{-1} x_i$ and each summand in the denominator by $x_j^{-1} x_j$, we obtain

$$x_i' = \frac{x_i^{-1} c_i^{-1}}{\sum_j x_j x_j^{-1} c_j^{-1}} x_i = \frac{f_i(\mathbf{x})}{\sum_j x_j f_j(\mathbf{x})} x_i,$$

i.e.

$$x_i' = \frac{f_i(\mathbf{x})}{\varphi(\mathbf{x})} x_i,$$

where

$$f_i(\mathbf{x}) = x_i^{-1} c_i^{-1} = \frac{1}{\sum_j a_{ij} x_i x_j}$$

supplies the *fitness* of grammar G_i in the population state $\mathbf{x}$, and

$$\varphi(\mathbf{x}) = \sum_j x_j f_j(\mathbf{x})$$

is average fitness. This shows that the system is again characterized by a replicator dynamic, albeit

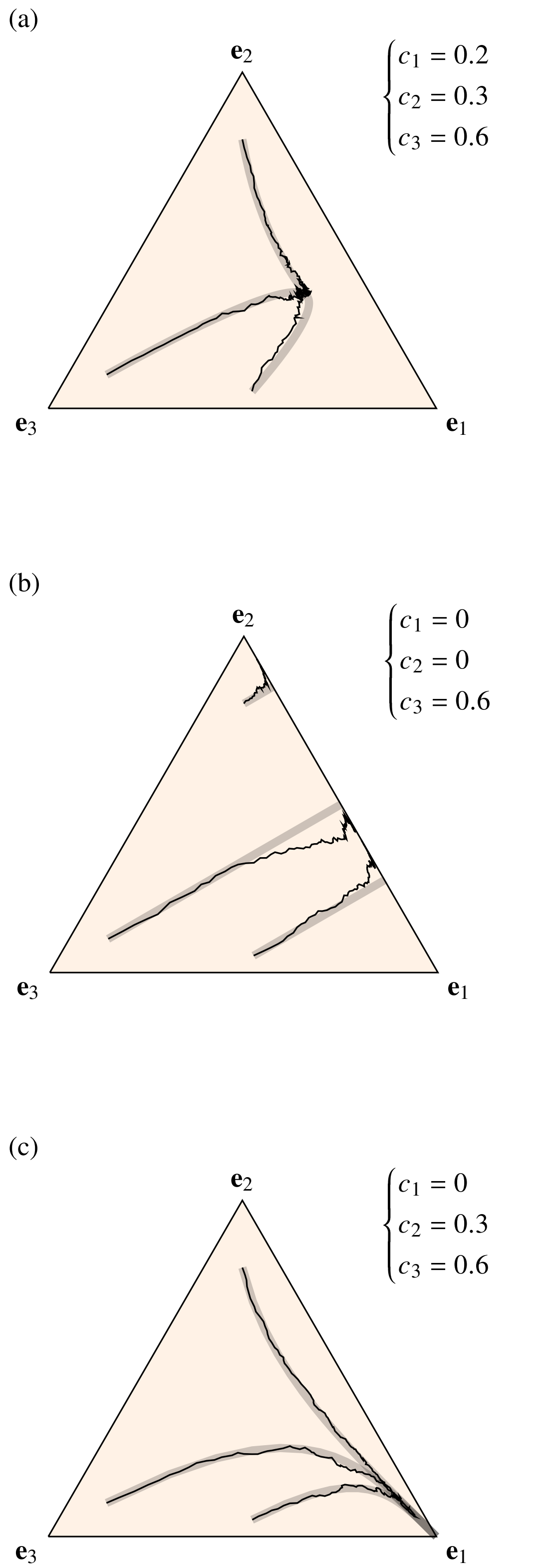

Figure 1: Learning in an omnipunitive (a) and two parapunitive (b, c) environments. Three learning trajectories are simulated in each case, starting from different initial states $\mathbf{W}$. In each case, the learning rate was set at $\gamma = 0.001$. The light thick lines show the corresponding mean dynamics.

in this more general case, fitness is frequency-dependent rather than constant. Note that the product $a_{ij}x_ix_j$ may be interpreted as the amount of motion away from grammar G_i in the direction of grammar G_j, explaining the appearance of these terms in the *denominator* of $f_i(\mathbf{x})$.

5 Global Asymptotic Dominance in n-grammar Competition

In the two-grammar case, Theorem 1 supplies a necessary and sufficient condition for one grammar to be an attractor. We now ask if a corresponding condition can be found in the more general case of n-grammar competition. The abundance vector $\mathbf{x}$, describing the abundances of the n grammars in a given generation of learners, belongs to the simplex $\Delta = \Delta^{n-1}$. At each vertex $\mathbf{e}_k \in \Delta$ of the simplex, one grammar claims all the abundance, and we define:

Definition 5. Grammar G_k is *dominant* if $\mathbf{e}_k$ is an equilibrium. It is *asymptotically dominant* if $\mathbf{e}_k$ is asymptotically stable. It is *globally asymptotically dominant*—abbreviated *g.a.d.*—if $\mathbf{e}_k$ is an attractor for any non-equilibrium initial state $\mathbf{x} \in \Delta$ with $x_k > 0$.

Clearly, Theorem 1 implies that, in the $n = 2$ case, G_k is g.a.d if and only if $a_{ik} > a_{ki}$. The restriction to initial states that satisfy $x_k > 0$ is a technicality which will be required in some of the proofs below; note that this assumption is innocuous since, empirically, convergence to a grammar can only occur after an innovation in the direction of this grammar has happened.

We moreover define:

Definition 6. Grammar G_i is *vulnerable* to grammar G_j if $a_{ij} > 0$. Grammar G_j is *non-submissive* if every other grammar G_i is vulnerable to it.

We then have the following result:

Theorem 3. *G_k is dominant only if it is non-submissive.*

Proof. Suppose G_k is not non-submissive. Then some G_i exists such that $a_{ik} = 0$. At the vertex $\mathbf{e}_k$, the penalty for this grammar G_i is

$$c_i = a_{ik}x_k = a_{ik} = 0.$$

Hence, G_k is not the only zero-penalty grammar, and so, by Theorem 2, $\overline{\mathbf{W}}$ does not generically converge to the vertex $\mathbf{e}_k$. In other words, in an environment in which only G_k is employed, a learner does not acquire full use of G_k. This implies that $\mathbf{e}_k$ is not an equilibrium and hence that G_k is not dominant. □

As an immediate consequence, we obtain that a non-submissive grammar cannot be globally asymptotically dominant:

Corollary 2. *G_k is g.a.d. only if it is non-submissive.*

In particular, the above results imply that no grammar G_i that stands in a subset relation to another grammar G_j, in the sense that $L_i \subseteq L_j$, can be (globally asymptotically) dominant.

Corollary 3. *Suppose $L_i \subseteq L_j$. Then G_i is not (globally asymptotically) dominant.*

Proof. $L_i \subseteq L_j$ implies that $a_{ji} = 0$. Hence G_j is not vulnerable to G_i and so G_i is not non-submissive. □

On the other hand, suppose a grammar G_k is non-submissive and that $a_{ki} = 0$ for all $i \neq k$. In this case, G_k is vulnerable to none of its competitors, while each of the latter is vulnerable to G_k. Let us call such a grammar *apical* (it is located at the "apex" of the competition situation).

Definition 7. Grammar G_k is *apical* if it is not vulnerable to any of its competitors but all its competitors are vulnerable to it.

An apical grammar is necessarily g.a.d.:

Theorem 4. *If G_k is apical, it is g.a.d.*

Proof. Let $\mathbf{x} \in \Delta$ such that $x_k > 0$. For all grammars G_i with $i \neq k$, the penalty is

$$c_i = \sum_j a_{ij}x_j \geq a_{ik}x_k > 0.$$

For G_k, the penalty is

$$c_k = \sum_j a_{kj}x_j = 0$$

on the assumption of apicality. Hence, by Theorem 2, G_k is g.a.d. (In fact, in this case, the vertex $\mathbf{e}_k$ is attained in one generation of learning.) □

We now continue to look for a general criterion for global asymptotic dominance. With the above results in mind, we can assume without loss of generality that a putative g.a.d. grammar G_k is both non-submissive (since if it were not non-submissive, it could not be g.a.d.) and non-apical (since if it were apical, it would necessarily be g.a.d.). On the assumption that G_k is non-submissive, we have $a_{ik} > 0$ for all $i \neq k$. Hence, whenever $x_k > 0$, we have

$$c_i = \sum_j a_{ij}x_j \geq a_{ik}x_k > 0.$$

On the other hand, since G_k is not apical, we have $a_{ki} > 0$ for at least one $i \neq k$, whereby

$$c_k = \sum_j a_{kj}x_j \geq a_{ki}x_i > 0$$

whenever $x_i > 0$. Hence, at least in the interior Δ°, every grammar has non-zero penalty, and so the learning environment is omnipunitive in all of the interior. Therefore, we have the well-posed population mean dynamic

$$x_i' = \frac{f_i(\mathbf{x})}{\varphi(\mathbf{x})}x_i$$

with

$$f_i(\mathbf{x}) = \frac{1}{\sum_j a_{ij}x_i x_j}$$

and $\varphi(\mathbf{x}) = \sum_j x_j f_j(\mathbf{x})$. By the following result, we can also assume without loss of any generality that the initial state $\mathbf{x}$ lies in the interior:

Lemma 1. *Suppose* $\mathbf{x} \in \partial\Delta$ *with* $x_k > 0$ *and that* G_k *is non-submissive. Then either* $\overline{\mathbf{W}}$ *converges to some point in the interior* Δ° *or it converges to the vertex* $\mathbf{e}_k$.

Proof. For any $i \neq k$, we have

$$c_i = \sum_j a_{ij}x_j \geq a_{ik}x_k > 0$$

on the assumption that G_k is non-submissive. Now, either $c_k > 0$ or $c_k = 0$. In the former case, every penalty probability is non-zero and so the learning environment is omnipunitive. By Theorem 2, $\overline{\mathbf{W}}$ converges to $\overline{\mathbf{W}}^* = c_i^{-1}/\sum_j c_j^{-1} \in \Delta^\circ$.

Then suppose $c_k = 0$. In this case, Corollary 1 implies that $\overline{\mathbf{W}}$ converges to the vertex $\mathbf{e}_k$. □

We now ask under what conditions trajectories starting at $\mathbf{x} \in \Delta^\circ$ converge to $\mathbf{e}_k$, so that G_k is g.a.d. (assuming, as above, that G_k is non-submissive and not apical). We define:

Definition 8. The *resilience* of grammar G_i against grammar G_j is $r_{ji} = 1/a_{ij}$, the inverse of the advantage that G_j holds over G_i. The *total resilience* of grammar G_i (against all its competitors) is defined as

$$R_i = \frac{1}{\sum_j a_{ij}},$$

the inverse of the cumulative advantage against G_i.

Of course, the resilience r_{ji} is only defined if $a_{ij} > 0$, i.e. only if G_i is vulnerable to G_j. In consequence, the total resilience R_i is only defined if G_i is vulnerable to at least one other G_j. Since a putative g.a.d. grammar G_k is assumed to be non-submissive and non-apical, the total resilience R_i is defined for all i, and the pairwise resiliences r_{ki} also exist for all i.

We point out that (total) resilience measures how robust a grammar G_i is against its competitor(s), i.e. it only references the advantage(s) of the latter over G_i, and not the advantage that G_i itself holds over its competitor(s).

We can now state our two main results. The first gives a sufficient condition for a grammar's global asymptotic dominance; the second, a necessary one. The proofs, which are tedious, are relegated to the Appendix.

Theorem 5. *Grammar* G_k *is g.a.d. if*

$$R_k > \sum_{i \neq k} r_{ki}$$

i.e. if its total resilience is greater than the sum of pairwise resiliences against it.

Theorem 6. *Grammar* G_k *is g.a.d. only if*

$$R_k > \frac{1}{n-1}\sum_{i \neq k} R_i$$

i.e. only if its total resilience is greater than the average total resilience across its competitors.

6 Grammar Flux

Theorem 5 enables an alternative characterization of sufficient conditions for a grammar's being g.a.d. This turns on the notion of a *chain* of grammars and its associated *flux*.

Definition 9. A *chain* of grammars is any triple (G_i, G_k, G_j) such that $i \neq k$ and $j \neq k$ (but possibly $i = j$). We denote this by $C_{ikj} = (G_i, G_k, G_j)$. A chain *traverses* grammar G_k if G_k is in the middle position of the chain. The *set of chains through* G_k is the set of all chains that traverse G_k.

Definition 10. The *flux* through a chain of grammars C_{ikj} is defined as the ratio

$$\varphi(C_{ikj}) = \frac{a_{kj}}{a_{ik}}$$

whenever $a_{ik} > 0$.

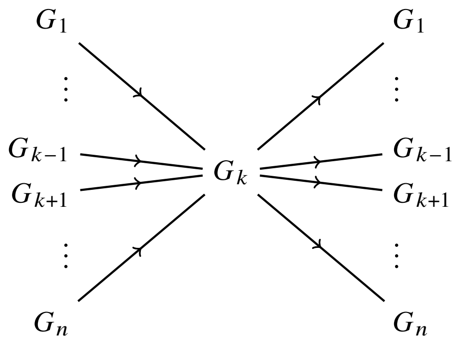

Figure 2: The total flux through grammar G_k, $\Phi(G_k)$, is defined by adding together the fluxes over all chains that traverse G_k.

If we can imagine abundance as "flowing" from one grammar to another, then $\varphi(C_{ikj})$ measures how much abundance, obtained from G_i, grammar G_k lets flow through to G_j. The better G_k is at resisting losing advantage to G_j under "pressure" from (advantage flowing from) G_i, the lower the flux.

We further define (cf. Figure 2):

Definition 11. The *total flux through grammar* G_k, denoted $\Phi(G_k)$, is the sum of fluxes over all chains through G_k, i.e. the quantity

$$\Phi(G_k) = \sum_{i \neq k} \sum_{j \neq k} \varphi(C_{ikj}).$$

A sufficient condition for g.a.d. is then the following: the total flux through a grammar is bounded from above by one.

Theorem 7. *Grammar G_k is g.a.d. if $\Phi(G_k) < 1$.*

Proof. By Theorem 5, G_k is g.a.d. if

$$\frac{1}{\sum_j a_{kj}} > \sum_{i \neq k} \frac{1}{a_{ik}}.$$

Since $a_{kk} = 0$, this is equivalent to

$$\frac{1}{\sum_{j \neq k} a_{kj}} > \sum_{i \neq k} \frac{1}{a_{ik}}.$$

Multiplying both sides by $\sum_{j \neq k} a_{kj}$, we obtain

$$1 > \sum_{j \neq k} a_{kj} \sum_{i \neq k} \frac{1}{a_{ik}} = \sum_{j \neq k} \sum_{i \neq k} \frac{a_{kj}}{a_{ik}} = \Phi(G_k)$$

as wished. □

A loose analogy exists between the notion of grammar flux and elementary properties of electronic circuits that may assist in the former's interpretation. By Ohm's Law, the current passed by a component, I, equals the voltage applied to the component, V, divided by its resistance, R. From this, one obtains $R^{-1} = I/V$, where R^{-1}, the inverse of resistance, measures the component's conductance, i.e. how many units of current it passes per unit of voltage applied. The flux through a grammar—even though a dimensionless number—similarly measures how much abundance a grammar "leaks" under the pressure of abundance flowing into it; Theorem 7 shows that, in order for a grammar to be globally asymptotically dominant, it is enough for this grammar to be a sufficiently poor conductor of abundance.

7 Discussion

We have studied the inter-generational dynamics of the general n-grammar variational learning model, asking the following question: under what conditions is a single grammar globally asymptotically dominant (g.a.d.), meaning that this grammar attracts from any non-equilibrium initial state once the grammar has been innovated? We have proved two main results, each turning on the notion of a grammar's resilience. The first result gives a sufficient condition, stating that grammar G_k is g.a.d. if its total resilience is greater than the cumulative pairwise resiliences of its competitors,

$$R_k > \sum_{i \neq k} r_{ki}.$$

The second result gives a necessary condition, and states that if G_k is g.a.d., then necessarily

$$R_k > \frac{1}{n-1} \sum_{i \neq k} R_i,$$

i.e. the total resilience of G_k is greater than the average total resilience computed over all its competitors. An alternative characterization by way of the notion of the flux through a grammar found that the sufficient condition is equivalent to $\Phi(G_k) < 1$, which states that the total flux through G_k is bounded from above by 1.

These results may now be applied to empirical data to evaluate the merits of this particular model of language acquisition and language change. The estimation of advantage parameters from corpora is by now commonplace in the two-grammar special case (e.g. Yang, 2000; Heycock and Wallenberg, 2013; Simonenko et al., 2019). In principle, estimation of such parameters in multidimensional competition is no different, although it is more

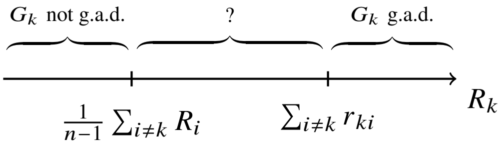

Figure 3: Necessary and sufficient bounds on R_k for grammar G_k to be globally asymptotically dominant.

laborious due to the greater number of pairwise comparisons required. The sufficient condition for global asymptotic dominance may then be used to predict whether a given set of estimates for the a_{ij} parameters (a particular advantage matrix A) leads to eventual dominance by one grammar. Conversely, the necessary condition may be used to argue that a particular choice of A could never lead to a grammar's being g.a.d. In practice, application of the sufficient condition for purposes of arguing for a grammar's g.a.d. status may be more useful than using the necessary condition to argue against a grammar's g.a.d. status—a particular advantage matrix may fail the necessary condition, and hence imply that a grammar is not formally g.a.d., yet the system's stable equilibrium may still be arbitrarily close to the vertex representing full use of this grammar.

The sufficient and necessary conditions of global asymptotic dominance reference the competing grammars' pairwise and total resiliences; in particular, both give a lower bound for the total resilience R_k of a candidate grammar. It is important to point out that the bounds for R_k established here do not coincide (except in the special case $n = 2$). In other words, a region exists in which mere examination of the magnitude of R_k does not (yet) tell us whether G_k is g.a.d. or not (see Figure 3). This is because the proofs of Theorems 5–6 rely on simple bounding arguments; in particular, the proof of the sufficient condition in fact establishes the stronger claim that convergence to dominance is strictly monotonic, in the sense that the abundance of the g.a.d. grammar always increases. If a trajectory converging on a vertex is non-monotonic, this is not captured by Theorem 5. Future work should thus look for a more complete characterization of the conditions under which convergence to the vertices occurs.

These results were obtained for a particular model of language change obtained from a particular model of language learning through a particular set of assumptions. We have assumed a discrete sequence of non-overlapping generations, each of which consists of infinitely many well-mixing, identically learning speakers. In other words, we have only studied the deterministic limit in which the system's mean dynamic is a faithful description of the system's evolution. Any of these simplifying assumptions could in principle be lifted, resulting in a stochastic process also at the inter-generational population level.

On the other hand, ample possibilities exist for future work even in the deterministic limit. We have here assumed, with tradition, that the pairwise advantages a_{ij} remain constant for the duration of any evolutionary process we may be interested in observing and modelling. This is not necessarily so in the real world. Interesting extensions of the model occur when the a_{ij} are allowed to be frequency-dependent, i.e. to depend on the current abundance vector $\mathbf{x}$ obtaining in the population. These await formal study.

Finally, we point out that the operators in (5) are but one out of many possible ways of generalizing the two-grammar learning algorithm for $n > 2$ grammars. In particular, this choice of operators implies that, whenever a grammar is punished, then *all* remaining $n - 1$ grammars are rewarded. Yet this is clearly not a necessary feature of a model of language learning, and may not be particularly realistic. An alternative model might equip the learner with a short-term memory, which would be used to reward only grammars employed by the learner in the very near past, for example. The effects such modifications have on the population-level evolution remain to be studied.

Acknowledgements

The work reported here has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 851423). I wish to thank Fernanda Barrientos, Caroline Heycock, Mara Maza-Cuello, Gemma McCarley, Raquel Montero Estebaranz, Molly Rolf, George Walkden, Joel Wallenberg and Charles Yang for discussions on the topic of grammar competition and variational learning; as well as the three reviewers for their helpful comments. I am particularly grateful to Mara Maza-Cuello for discovering and helping to rectify an embarrassing mistake in an earlier version of the proof of Theorem 6. Any remaining errors, of course, are my sole responsibility.

References

Henning Andersen. 1973. Abductive and deductive change. *Language*, 49(4):765–793.

Robert R. Bush and Frederick Mosteller. 1955. *Stochastic models for learning*. Wiley, New York, NY.

Caroline Heycock and Joel Wallenberg. 2013. How variational acquisition drives syntactic change: the loss of verb movement in Scandinavian. *Journal of Comparative Germanic Linguistics*, 16:127–157.

Henri Kauhanen. 2019. Stable variation in multidimensional competition. In Anne Breitbarth, Miriam Bouzouita, Lieven Danckaert, and Melissa Farasyn, editors, *The Determinants of Diachronic Stability*, pages 263–290. Benjamins, Amsterdam.

Anthony S. Kroch. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1(3):199–244.

David J. C. MacKay. 2003. *Information theory, inference, and learning algorithms*. Cambridge University Press, Cambridge.

Kumpati S. Narendra and Mandayam A. L. Thathachar. 1989. *Learning automata: an introduction*. Prentice-Hall, Englewood Hills, NJ.

Partha Niyogi. 2006. *The computational nature of language learning and evolution*. MIT Press, Cambridge, MA.

M. Frank Norman. 1972. *Markov processes and learning models*. Academic Press, New York, NY.

William H. Sandholm. 2010. *Population games and evolutionary dynamics*. MIT Press, Cambridge, MA.

Alexandra Simonenko, Benoit Crabbé, and Sophie Prévost. 2019. Agreement syncretization and the loss of null subjects: quantificational models for Medieval French. *Language Variation and Change*, 31(3):275–301.

Uriel Weinreich, William Labov, and Marvin I. Herzog. 1968. Empirical foundations for a theory of language change. In Winfred P. Lehmann and Yakov Malkiel, editors, *Directions for historical linguistics: a symposium*, pages 95–195. University of Texas Press, Austin, TX.

Charles D. Yang. 2000. Internal and external forces in language change. *Language Variation and Change*, 12:231–250.

Charles D. Yang. 2002. *Knowledge and learning in natural language*. Oxford University Press, Oxford.

A Proofs

Proof of Theorem 5. Let $\mathbf{x} \in \Delta^\circ$. We look for a condition under which $x'_k > x_k$. If this holds for arbitrary $\mathbf{x} \in \Delta^\circ$, then x_k increases in the entire interior of the simplex, which suffices to show that G_k is g.a.d.

Now, $x'_k > x_k$ if

$$f_k(\mathbf{x}) > \sum_i x_i f_i(\mathbf{x}).$$

Write $x_k = 1 - \delta$ for $\delta > 0$. Then the above inequality becomes

$$f_k(\mathbf{x}) > (1-\delta) f_k(\mathbf{x}) + \sum_{i \neq k} x_i f_i(\mathbf{x}),$$

or

$$\delta f_k(\mathbf{x}) > \sum_{i \neq k} x_i f_i(\mathbf{x}).$$

Expanding the fitnesses, we have

$$\frac{\delta}{\sum_i a_{ki}(1-\delta)x_i} > \sum_{i \neq k} \frac{x_i}{\sum_j a_{ij} x_i x_j},$$

i.e.

$$\frac{\delta}{\sum_i a_{ki} x_i} > (1-\delta) \sum_{i \neq k} \frac{1}{\sum_j a_{ij} x_j}. \tag{7}$$

Since $a_{kk} = 0$, this is equivalent to

$$\frac{\delta}{\sum_{i \neq k} a_{ki} x_i} > (1-\delta) \sum_{i \neq k} \frac{1}{\sum_j a_{ij} x_j}. \tag{8}$$

Since $x_k = 1 - \delta$, it follows that $x_i \leq \delta$ for $i \neq k$. Thus, the left-hand side is bounded from below; in fact, it is bounded from below by R_k:

$$\frac{\delta}{\sum_{i \neq k} a_{ki} x_i} \geq \frac{\delta}{\sum_{i \neq k} a_{ki} \delta} = \frac{1}{\sum_{i \neq k} a_{ki}} = R_k.$$

Hence, to establish (8), it suffices to show that

$$R_k > (1-\delta) \sum_{i \neq k} \frac{1}{\sum_j a_{ij} x_j}.$$

Working on the right-hand side, we find

$$\sum_{i \neq k} \frac{1-\delta}{\sum_j a_{ij} x_j} = \sum_{i \neq k} \frac{1-\delta}{a_{ik}(1-\delta) + \sum_{j \neq k} a_{ij} x_j} = \sum_{i \neq k} \frac{1}{a_{ik} + \frac{1}{1-\delta} \sum_{j \neq k} a_{ij} x_j},$$

which is obviously bounded from above:

$$\sum_{i \neq k} \frac{1}{a_{ik} + \frac{1}{1-\delta} \sum_{j \neq k} a_{ij} x_j} < \sum_{i \neq k} \frac{1}{a_{ik}} = \sum_{i \neq k} r_{ki}.$$

Hence, the condition

$$R_k > \sum_{i \neq k} r_{ki}$$

implies that $x'_k > x_k$ everywhere in the interior of Δ, which implies that G_k is g.a.d. □

Proof of Theorem 6. We in fact show the following stronger result: if $R_k < \sum_{i \neq k} R_i/(n-1)$, the vertex $\mathbf{e}_k$ is not asymptotically stable. From this it immediately follows that G_k is not g.a.d.

To prove that $\mathbf{e}_k$ is not asymptotically stable, we show that, in (possibly infinitesimally small) local environments of the vertex, $x'_k < x_k$, so that $\mathbf{x}$ moves away from $\mathbf{e}_k$. Let $\mathbf{x} \in \Delta^\circ$ such that $x_k = 1 - \delta$ for $\delta > 0$. The condition $x'_k < x_k$ is equivalent to

$$\frac{\delta}{1-\delta} \cdot \frac{1}{\sum_i a_{ki} x_i} < \sum_{j \neq k} \frac{1}{\sum_i a_{ji} x_i}. \tag{9}$$

(cf. the derivation of (7) in the proof of Theorem 5). Let $m = \min_{i \neq k}\{x_i\}$. Then $x_i \geq m$ for all $i \neq k$, and so

$$\frac{1}{\sum_i a_{ki} x_i} \leq \frac{1}{m \sum_i a_{ki}}.$$

Hence, to show (9), it suffices to show that

$$\frac{\delta}{1-\delta} \cdot \frac{1}{m} \cdot \frac{1}{\sum_i a_{ki}} < \sum_{j \neq k} \frac{1}{\sum_i a_{ji} x_i}. \tag{10}$$

Suppose δ is small ($\mathbf{x}$ is close to $\mathbf{e}_k$). Then $x_i < 1 - \delta$ for all $i \neq k$. We have

$$\sum_i a_{ji} x_i = a_{jk}(1-\delta) + \sum_{i \neq k} a_{ji} x_i,$$

and so (10) is equivalent to

$$\frac{\delta}{m} \cdot \frac{1}{\sum_i a_{ki}} < \sum_{j \neq k} \frac{1}{a_{jk} + \frac{1}{1-\delta} \sum_{i \neq k} a_{ji} x_i}. \tag{11}$$

Since $x_i < 1 - \delta$ (for $i \neq k$), the right-hand side is bounded from below by

$$\sum_{j \neq k} \frac{1}{a_{jk} + \frac{1}{1-\delta} \sum_{i \neq k} a_{ji}(1-\delta)} = \sum_{j \neq k} \frac{1}{\sum_i a_{ji}}.$$

Hence, to show (11), it suffices to show that

$$\frac{\delta}{m} \cdot \frac{1}{\sum_i a_{ki}} < \sum_{j \neq k} \frac{1}{\sum_i a_{ji}}$$

i.e. that

$$\frac{1}{\sum_i a_{ki}} < \frac{m}{\delta} \sum_{j \neq k} \frac{1}{\sum_i a_{ji}},$$

i.e.

$$R_k < \frac{m}{\delta} \sum_{j \neq k} R_j.$$

Let us rewrite this last inequality:

$$1 - \delta > 1 - m\rho_k,$$

where $\rho_k = \sum_{j \neq k} R_j / R_k$. Since $x_k = 1 - \delta$, we thus have

$$x_k > 1 - m\rho_k.$$

Now, since $\mathbf{x} \in \Delta$ and since $m = \min_{i \neq k}\{x_i\}$, we have

$$\begin{aligned} 1 &= x_k + \sum_{i \neq k} x_i \\ &\geq x_k + (n-1)m \\ &> 1 - m\rho_k + (n-1)m, \end{aligned}$$

which is equivalent to

$$\rho_k > n - 1.$$

In other words,

$$\sum_{j \neq k} \frac{R_j}{R_k} > n - 1,$$

or

$$R_k < \frac{1}{n-1} \sum_{j \neq k} R_j. \tag{12}$$

To recap: if condition (12) holds, then some neighbourhood of $\mathbf{e}_k$ exists in Δ° in which the value of x_k increases. This implies that $\mathbf{e}_k$ is not asymptotically stable; *a fortiori*, G_k is not g.a.d. □