fulness) constraints work against (towards) the preservation of the underlying contrasts. Thus, a small (large) language should arise by ranking the markedness (faithfulness) constraints as high as possible. And a ranking that ranks the markedness (faithfulness) constraints as high (low) as possible is a ranking that minimizes Prince and Tesar's function (9).

I endorse Prince and Tesar's conjecture that (9) is a restrictiveness measure, at least for the cases of interest.[1] In Magri (2012a), I backup this claim by looking at a case study, namely the typology corresponding to the large constraint set considered in Pater and Barlow (2003). In the rest of this paper, I thus focus on the reformulation (7) of the problem of the acquisition of phonotactics, with $\mu$ defined as in (9). The latter formulation of the problem of the acquisition of phonotactics is NP-complete too (Magri, 2010; Magri, 2012b). In the rest of this paper, I thus develop an integer programming formulation of the latter problem, that allows approximation algorithms for integer programming to be used in order to tackle the problem of the acquisition of phonotactics. The reasoning is split up into two steps. In Section, 2, I develop an integer programming formulation of the objective function, namely the alleged restrictiveness measure in (9). And in Section 3, I turn to an integer programming formulation of the OT-consistency condition.

## 2 An integer programming restatement of the restrictiveness measure

A *square matrix of order* $n$ is a collection of $n^2$ real numbers displayed into $n$ columns and $n$ rows. I denote a square matrix of order $n$ as $\mathbf{X} = [x_{i,j}]_{i,j=1}^n$, with the understanding that $x_{i,j}$ is the element of the matrix $\mathbf{X}$ which sits in the $i$th row and the $j$th column. I denote by $\mathbb{R}^{n\times n}$ the vector space of all square matrices of order $n$.

A square matrix $\mathbf{X} = [x_{i,j}]_{i,j=1}^n$ is called a *permutation matrix* iff its elements $x_{i,j}$ satisfy the following three conditions: (i) they are all 0 or 1; (ii) each column contains a unique 1; (iii) each row contains a unique 1. I denote by $\mathcal{P}^n$ the set of all $n!$ permutation matrices of order $n$. To illustrate, I list $\mathcal{P}^n$ with $n = 3$ in (10).

(10)
$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}$$
$$\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}\begin{bmatrix}0&1&0\\0&0&1\\1&0&0\end{bmatrix}\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix}$$

Permutation matrices play a special role in convex geometry (Webster, 1984, par. 5.8).

There is a natural correspondence between permutation matrices of order $n$ and rankings over $n$ constraints $C_1, \dots, C_n$. Recall that a ranking $\pi$ is a permutation over $\{1, 2, \dots, n\}$, with the understanding that $\pi(i) = j$ means that the ranking $\pi$ assigns the constraint $C_i$ to the $j$th stratum, with the convention that the stratum corresponding to $j = n$ is the top stratum. I use $i$ as the index ranging over constraints and $j$ as the index ranging over strata. Thus, a ranking $\pi$ can be identified with that (unique) permutation matrix $\mathbf{X} = [x_{i,j}]_{i,j=1}^n \in \mathcal{P}^n$ such that $x_{i,j} = 1$ iff the ranking $\pi$ assigns the constraint $C_i$ to the $j$th stratum, namely $\pi(i) = j$. To illustrate, I list in (11) the rankings over $\{C_1, C_2, C_3\}$ corresponding to the six permutation matrices in (10), respectively.

(11)
$$\begin{array}{lll} C_3 \gg C_2 \gg C_1, & C_2 \gg C_3 \gg C_1, & C_3 \gg C_1 \gg C_2, \\ C_1 \gg C_3 \gg C_2, & C_2 \gg C_1 \gg C_3, & C_1 \gg C_2 \gg C_3 \end{array}$$

I denote by $\pi_{\mathbf{X}}$ the ranking corresponding to a permutation matrix $\mathbf{X} \in \mathcal{P}^n$ and by $\mathbf{X}_\pi \in \mathcal{P}^n$ the permutation matrix corresponding to a ranking $\pi$. Prince and Tesar's restrictiveness measure (9) of a ranking $\pi$ can be straightforwardly read off the corresponding permutation matrix $\mathbf{X}_\pi$, as follows.

Define the *scalar product* $\langle \mathbf{X}, \mathbf{Y} \rangle \in \mathbb{R}$ between two arbitrary square matrices $\mathbf{X} = [x_{i,j}]_{i,j=1}^n, \mathbf{Y} = [y_{i,j}]_{i,j=1}^n \in \mathbb{R}^{n\times n}$ as in (12) (namely as the Euclidean scalar product of $\mathbb{R}^{n^2}$).

[1] Prince and Tesar's conjecture that (9) is a restrictiveness measure runs into a straightforward problem when the constraint set $\mathfrak{C}$ contains both positional and faithfulness constraints. Yet, there are various ways to circumvent this difficulty posed by positional constraints. One way could be to weigh differently the two types of faithfulness constraints in the determination of restrictiveness. Thus, we could switch from the definition in (9) to the variant in (i), where $\mathcal{F}_{pos}$ is the set of positional faithfulness constraints, $\mathcal{F}_{gen}$ is the set of general faithfulness constraints and $\alpha$ is a positive coefficient.

(i)
$$\mu_\alpha(\pi) \doteq \sum_{F\in\mathcal{F}_{pos}} \left|\left\{M \in \mathcal{M} \,\middle|\, \pi(F) > \pi(M)\right\}\right| + \alpha \sum_{F\in\mathcal{F}_{gen}} \left|\left\{M \in \mathcal{M} \,\middle|\, \pi(F) > \pi(M)\right\}\right|$$

Another way to deal with positional faithfulness constraints could be to ignore altogether rankings where a positional faithfulness constraint is ranked below the corresponding general faithfulness constraint. This is trivial to obtain, by adding a proper ERC to the ERC matrix given with an instance of the problem (7).

(12) $$\langle \mathbf{X}, \mathbf{Y} \rangle \doteq \sum_{i,j=1}^{n} x_{i,j} y_{i,j}.$$

A function $f : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ is called *linear* iff there exists a square matrix $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ such that (13) holds for any square matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$.

(13) $$f(\mathbf{X}) = \langle \mathbf{\Sigma}, \mathbf{X} \rangle.$$

Linear functions are the "simplest" possible convex functions, namely the ones that yield the easiest optimization problems.

Let me assume that the first $m$ constraints in $\mathfrak{C}$ are the faithfulness constraints while the remaining $n - m$ constraints are the markedness constraints, namely that $\mathcal{F} = \{C_1, \ldots, C_m\}$ and $\mathcal{M} = \{C_{m+1}, \ldots, C_n\}$. Consider the matrix $\mathbf{\Sigma}_{n,m} \in \mathbb{R}^{n \times n}$ defined as follows: its first $m$ rows each have the form $[0, 1, \ldots, n-2, n-1]$; the remaining $n - m$ rows are all null. To illustrate, I give in (14) the matrix $\mathbf{\Sigma}_{n,m}$ with $n = 7, m = 4$.

(14) $$\mathbf{\Sigma}_{7,4} \doteq \left[\begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The following Claim 1 explains how to compute the restrictiveness $\mu(\pi)$ of a ranking $\pi$ according to (9) out of the corresponding permutation matrix $\mathbf{X}_\pi$; see Appendix A.1. This Claim shows an important property of Prince and Tesar's restrictiveness measure: it can be described as a linear function over the set of permutation matrices.

**Claim 1** *The restrictiveness $\mu(\pi)$ of a ranking $\pi$ according to (9) can be computed as follows:*

*(15)* $$\mu(\pi) = \langle \mathbf{\Sigma}_{n,m}, \mathbf{X}_\pi \rangle - \frac{1}{2} m(m-1)$$

*namely as the scalar product $\langle \mathbf{\Sigma}_{n,m}, \mathbf{X} \rangle$ between the matrix $\mathbf{\Sigma}_{n,m}$ and the corresponding permutation matrix $\mathbf{X}_\pi$, minus the constant $\frac{1}{2}m(m-1)$ which does not depend on the ranking.*[2] ■

[2]I have noted in footnote 1 that the conjecture that the function $\mu$ in (9) is a restrictiveness measure runs into problems for constraint sets that contain both general and positional faithfulness constraints. And I have suggested that a possible way out is to to switch from the definition (9) to the variant in (i). Let me now point out that the latter variant too can be described as a linear function over permutation matrices. In fact, let $\mathbf{\Sigma}_{n,m,\alpha}$ be as the matrix $\mathbf{\Sigma}_{n,m}$ defined above, but with the rows corresponding to general faithfulness constraints multiplied by $\alpha$. Then, $\mu_\alpha(\mathbf{X})$ coincides with $\langle \mathbf{\Sigma}_{n,m,\alpha}, \mathbf{X} \rangle$, but for a constant.

The problem of the acquisition of phonotactics (7) with Prince and Tesar's alleged restrictiveness measure (9) can thus be restated as the optimization problem (16).

(16) *minimize*: $\langle \mathbf{\Sigma}_{n,m}, \mathbf{X} \rangle$;
*subject to*: $\mathbf{X} \in \mathcal{P}^n$ and $\pi_{\mathbf{X}}$ is consistent with the given ERC matrix $\mathbf{E}$.

Here, I have dropped the constant $\frac{1}{2}m(m-1)$ which appears in (15), as it does not affect the optimization problem.

## 3 An integer programming formulation of the OT-consistency condition

The reformulation in (16) makes use of the notion of OT-consistency with a given ERC matrix and this notion is currently stated in terms of rankings rather than in terms of the corresponding permutation matrices. We need to restate the latter condition directly in terms of permutation matrices. In this Section, I point out two strategies for doing that. The first approach hinges on a classical observation by Prince and Smolensky (2004) that OT consistency can be restated as linear consistency in the case of exponentially spaced weights. The second approach requires a larger number of linear conditions, but is shown to provide a better reformulation (i.e. a tighter relaxation).

### 3.1 An initial formulation of OT-consistency

Given an ERC $\mathbf{e} = [e_1, \ldots, e_n]$, consider the corresponding square matrix $\mathbf{A_e} = [a_{i,j}]_{i,j=1}^n \in \mathbb{R}^{n \times n}$ defined in (17). Here, $t_i$ is the *sign* of the ERC's entry $e_i$, namely $t_i$ is equal to $-1$, $0$ or $+1$ depending on whether $e_i$ is equal to L, $e$ or W. Thus, the entry $a_{i,j}$ in the $i$th row and the $j$th column of the matrix (17) consists of the sign $t_i$ multiplied by $2^j$.

(17) $$\mathbf{A_e} = \begin{bmatrix} 2^1 t_1 & 2^2 t_1 & \ldots & 2^j t_1 & \ldots & 2^n t_1 \\ & & & \vdots & & \\ 2^1 t_i & 2^2 t_i & \ldots & 2^j t_i & \ldots & 2^n t_i \\ & & & \vdots & & \\ 2^1 t_n & 2^2 t_n & \ldots & 2^j t_n & \ldots & 2^n t_n \end{bmatrix}$$

Intuitively, this entry $a_{i,j} = 2^j t_i$ is the weight of the sign $t_i$ under the assumption that the constraint

$C_i$ is assigned to the $j$th stratum.

The following claim offers a restatement of OT-consistency between an ERC and a ranking in terms of the permutation matrix corresponding to that ranking. This claim is just a restatement in matrix form of the observation by Prince and Smolensky (2004) that OT consistency is equivalent to a linear condition with exponentially spaced weights; see Subsection A.2.

**Claim 2** *A ranking $\pi$ is OT-consistent with an ERC $\mathbf{e}$ iff $\langle \mathbf{A}_{\mathbf{e}}, \mathbf{X}_\pi \rangle \geq 0$, where $\langle \mathbf{A}_{\mathbf{e}}, \mathbf{X}_\pi \rangle$ is the scalar product (12) between the matrix $\mathbf{A}_{\mathbf{e}}$ corresponding to the ERC $\mathbf{e}$ and the permutation matrix $\mathbf{X}_\pi$ corresponding to the ranking $\pi$.* ■

The current formulation (16) of the problem of the acquisition of phonotactics can thus be restated as the optimization problem in (18).

(18) FIRST INTEGER REFORMULATION:

*minimize*: $\langle \mathbf{\Sigma}_{n,m}, \mathbf{X} \rangle$;

*subject to*: $\mathbf{X} \in \mathcal{P}^n$ s.t. $\langle \mathbf{A}_{\mathbf{e}}, \mathbf{X} \rangle \geq 0$ for every ERC $\mathbf{e}$ of the ERC matrix $\mathbf{E}$.

Problem (18) is an optimization problem over permutation matrices $\mathbf{X} \in \mathcal{P}^n$. The *objective function* is the linear function $\langle \mathbf{\Sigma}_{n,m}, \mathbf{X} \rangle$. And the feasible set is defined in terms of linear *side conditions* $\langle \mathbf{A}_{\mathbf{e}}, \mathbf{X} \rangle \geq 0$. Problem (18) is thus an *integer program*. In particular, it is an *Assignment problem* with *linear side constraints* (AssignLSCsPbm) (Bertsimas and Weismantel, 2005).

### 3.2 Another formulation of OT-consistency

Let $\ell(\mathbf{e})$ be the number of entries equal to L in an ERC $\mathbf{e} = [e_1, \ldots, e_n]$. Assume without loss of generality that $\ell(\mathbf{e}) > 0$, as ERCs with no L's can be ignored. For every stratum $\bar{\jmath} \in \{1, \ldots, n\}$, consider the square matrix $\mathbf{A}_{\mathbf{e}}^{\bar{\jmath}} = [a_{i,j}]_{i,j=1}^n$ with $n$ rows and $n$ columns whose generic element $a_{i,j}$ is defined as in (19).

(19) $$a_{i,j} \doteq \begin{cases} 1 & \text{if } e_i = \text{L},\ j \geq \bar{\jmath} \\ -1 & \text{if } e_i = \text{W},\ j \geq \bar{\jmath} + \ell \\ 0 & \text{otherwise} \end{cases}$$

The following claim offers another restatement of OT-consistency between an ERC and a ranking in terms of the permutation matrix corresponding to that ranking; see Subsection A.3.

**Claim 3** *A ranking $\pi$ is OT-consistent with an ERC $\mathbf{e}$ iff $\langle \mathbf{A}_{\mathbf{e}}^{\bar{\jmath}}, \mathbf{X}_\pi \rangle \leq 0$ for every $\bar{\jmath} \in \{1, \ldots, n\}$, where $\langle \mathbf{A}_{\mathbf{e}}^{\bar{\jmath}}, \mathbf{X}_\pi \rangle$ is the scalar product (12) between the matrix $\mathbf{A}_{\mathbf{e}}^{\bar{\jmath}}$ corresponding to the ERC $\mathbf{e}$ and the stratum $\bar{\jmath}$ and the permutation matrix $\mathbf{X}_\pi$ corresponding to the ranking $\pi$.* ■

The current formulation (16) of the problem of the acquisition of phonotactics can thus be alternatively restated as the optimization problem (20).

(20) SECOND INTEGER REFORMULATION:

*minimize*: $\langle \mathbf{\Sigma}_{n,m}, \mathbf{X} \rangle$;

*subject to*: $\mathbf{X} \in \mathcal{P}^n$ s.t. $\langle \mathbf{A}_{\mathbf{e}}^{\bar{\jmath}}, \mathbf{X} \rangle \leq 0$ for every ERC $\mathbf{e}$ of the ERC matrix $\mathbf{E}$ and every $\bar{\jmath} \in \{1, \ldots, n\}$.

Again, (20) is another instance of the AssignLSCsPbm. The feasible set in the latter formulation (20) involves $n$ times more inequalities than the formulation (18).

### 3.3 Comparing the two formulations

Problems (18) and (20) are two different formulations of the original problem (16) of the acquisition of phonotactics. They are thus equivalent, in the sense that a solution to any of the two problems is also a solution to the other and furthermore to the original problem. This Subsection explains why, nonetheless, the latter formulation (20) is better than the former formulation (18).

Both (18) and (20) are optimization problems over permutation matrices $\mathbf{X} \in \mathcal{P}^n$. The latter condition on the matrix $\mathbf{X} = [x_{i,j}]_{i,j=1}^n$ means that conditions (21) hold for any $i, j = 1, \ldots, n$.

(21) $$x_{i,j} \in \{0,1\}$$ $$\sum_{i=1}^n x_{i,j} = 1, \quad \sum_{j=1}^n x_{i,j} = 1$$

Problems (18) and (20) are *integer* optimization problems because of the condition $x_{i,j} \in \{0,1\}$ in (21). This condition can be *relaxed*, requiring the entires $x_{i,j}$ to be not necessarily 0 or 1 but instead any number in between 0 and 1. Thus, let $\mathcal{P}_{\text{rel}}^n$ be the set of matrices that satisfy the relaxed conditions (22), known as the *Birkhoff polytope*.

(22) $$x_{i,j} \in [0,1]$$ $$\sum_{i=1}^n x_{i,j} = 1, \quad \sum_{j=1}^n x_{i,j} = 1$$

Relaxing the integer constraint $\mathbf{X} \in \mathcal{P}^n$ into the continuous constraint $\mathbf{X} \in \mathcal{P}_{\text{rel}}^n$, yields the two corresponding problems (23) and (24).