The error-driven ranking model of the early stage of the acquisition of phonotactics: An initial result on restrictiveness*

Giorgio Magri

SFL (CNRS, University of Paris 8)

Nine-month-old infants are already sensitive to the distinction between licit and illicit forms (Jusczyk et al. 1993). They thus display knowledge of the target adult phonotactics at an early stage when morphology is plausibly still lagging behind (Hayes 2004) and the acquisition of the native language lexicon has barely begun (Fenson et al. 1994). How can this *early stage* of the acquisition of phonotactics be modeled? According to the *error-driven* learning model, the child is trained on a stream of licit forms, starts from a grammar which corresponds to the most restrictive phonotactics, and slightly relaxes its current grammar whenever it fails on the current piece of data. The model does not require a lexicon, as the slight updates of the current grammar are based on a single piece of data at the time. Furthermore, the model is trained only on licit forms and thus does not require alternations, which might still be unavailable at a stage when morphology is lagging behind. Finally, the model predicts a sequence of grammars that can be matched with child acquisition paths, thus modeling the child's acquisition gradualness. Section 1 describes an implementation of the error-driven model within Optimality Theory (OT; Prince and Smolensky 2004).

Being trained on licit forms, it is easy to guarantee that the phonotactics learned by the model is *consistent*, namely that it correctly classifies as licit any form which is indeed licit according to the target phonotactics. Yet, the phonotactics learned by the model could fail at *restrictiveness*: it could incorrectly classify as licit also forms which are instead illicit according to the target phonotactics. Restrictiveness is indeed the main issue of the computational theory of the error-driven model of the early stage of the acquisition of phonotactics, as discussed in section 2.

Prince and Tesar (2004) and Hayes (2004) have identified the challenge raised by restrictiveness in the problem of learning a restrictive relative ranking of the faithfulness constraints. Yet, we intuitively expect the relative ranking of the faithfulness constraints to be crucial for the phonology (namely for the specific way in which illicit forms are repaired), but less crucial for phonotactics (namely for the divide between licit and illicit forms). Section 3 formalizes the intuitive condition that the relative ranking of the faithfulness constraints does not matter to describe a certain phonotactic pattern. These con-

*The research reported here was supported in part by a grant from the Fyssen Research Foundation.

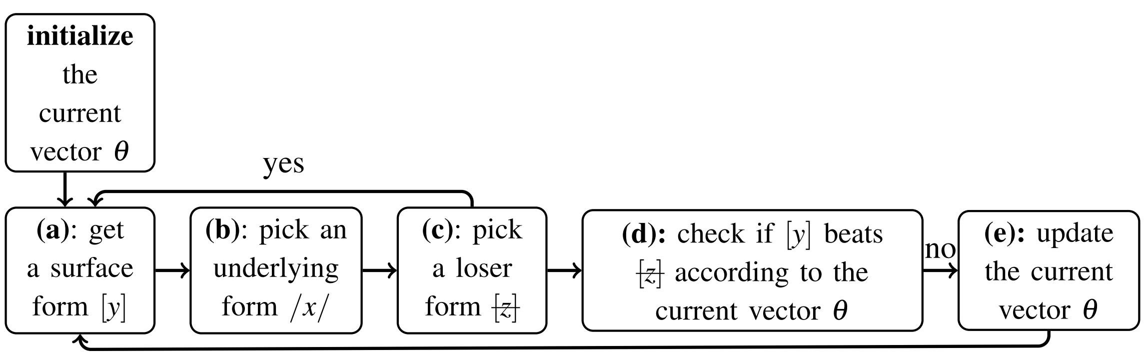
siderations motivate the main question addressed in this paper: if the target phonotactics does not require any specific relative ranking of the faithfulness constraints, is it possible to provide guarantees that the OT error-driven learning model is restrictive, and thus succeeds at learning the target phonotactics?

The main result of this paper consists of theorem 3 in section 4, which provides a positive answer to this question. The crucial assumption of the theorem is that the faithfulness constraints don't raise too high. Under this assumption, the OT error-driven model is guaranteed to be restrictive when trained on a phonotactic pattern which does not require any specific relative ranking of the faithfulness constraints. These guarantees hold under extremely mild assumptions on the constraint set (see assumptions 1, 2, and 3 below). In other words, this result only relies on the intrinsic ranking logic of OT, not on the additional structure that would be introduced through restrictive assumptions on the constraint set. Section 5 provides a sketch of the proof of the theorem and section 6 frames the result within the larger project it belongs to, as articulated in Magri (2013a,e,c,b,d).

1. The EDRA model of the acquisition of phonotactics

An OT typology is parameterized by the set of all possible rankings over a given, finite set of phonological constraints $C_1, \ldots, C_n$. As noted in Boersma (1997, 1998), rankings can be represented numerically, as follows. Each constraint C_k is assigned a numerical *ranking value* θ_k. These ranking values are collected into an n-tuple $\theta = (\theta_1, \ldots, \theta_n)$, called a *ranking vector*. The relative size of the ranking values naturally defines a rankings over the constraint set: a constraint C_k is ranked above another constraint C_h whenever the ranking value θ_k of the former is larger than or equal to the ranking value θ_h of the latter. If some of the ranking values tie, then the ranking vector represents different rankings, that differ in how they break the ties. The error-driven learning scheme within OT can then be described as the algorithm in (1).

(1) *The OT error-driven ranking algorithm*



The algorithm maintains a current hypothesis of the target OT grammar, represented through a current vector $\theta = (\theta_1, \ldots, \theta_n)$ of ranking values. These current ranking values are initialized and then updated by looping though the five steps (1a)-(1e) described below.

At steps (1a)-(1c), the algorithm assembles a triplet (/x/, [y], ~~[z]~~) of an underlying form /x/ and two candidates [y] and ~~[z]~~ together with the information that the former is the winner and the latter the loser according to the target OT grammar the algorithm is being trained on (as a mnemonic, I strike out a candidate when it is meant to be a loser). At step (1d), the algorithm checks whether its current hypothesis of the target OT grammar manages to make the intended winner [*y*] beat the intended loser ~~[*z*]~~. In terms of a ranking $\gg$, this condition requires that the top $\gg$-ranked among the *winner-preferring constraints* (namely those constraints that assign *less* violations to the winner [*y*] than to the loser ~~[*z*]~~) is $\gg$-ranked above the top $\gg$-ranked among the *loser-preferring constraints* (namely those constraints that assign *more* violations to the winner [*y*] than to the loser ~~[*z*]~~). Translated in terms of a ranking vector, this condition says that the largest ranking value over winner-preferring constraints is larger than the ranking value of any loser-preferring constraint.

If indeed the current ranking values manage to make the winner [*y*] beat the loser candidate ~~[*z*]~~, then the algorithm has nothing to learn from this comparison, it loops back to step (1a) and waits for more data. If instead the algorithm fails, then it needs to slightly revise the current ranking values at step (1e). The current failure suggests that some or all of the ranking values of the loser-preferring (or winner-preferring) constraints are too large (too small, respectively). The algorithm thus promotes the winner-preferring constraints by a small *promotion amount* and demotes the loser-preferring constraints by a small *demotion amount*. What matters is not the actual values of the promotion and demotion amounts, but rather their ratio. Thus, the demotion amount can be set equal to 1 for concreteness, letting instead the promotion amount be equal to an arbitrary non-negative constant p, as in (2).

(2) a. Increase the ranking value of each winner-preferring constraint by $p \geq 0$;
b. decrease the ranking value of each *undominated* loser-preferring constraint by 1.

Crucially, the demotion component (2b) of the re-ranking rule does not demote all loser-preferring constraints, but only those that really need to be demoted, namely those loser-preferring constraints which are not currently ranked underneath a winner-preferring constraint and are therefore called *undominated* (Tesar and Smolensky 1998). Translated in terms of a ranking vector $\theta = (\theta_1, \ldots, \theta_n)$, the latter condition says that a loser-preferring constraint C_k is undominated provided its ranking value θ_k is at least as large as the largest ranking value over winner-preferring constraints. The algorithm (1) with the re-ranking rule (2) is called an *error-driven ranking algorithm* (EDRA).

An EDRA can be turned into a model of the child's early stage of the acquisition of phonotactics through a proper choice of the implementations details, as follows. The markedness constraints start out initially ranked above the faithfulness constraints (Smolensky 1996). Concretely, I assume that the faithfulness constraints start out with a null initial ranking value while the markedness constraints start out with an initial ranking value equal to a large constant $\theta_{\text{init}} > 0$. These initial ranking values are then updated through the five steps (1). I assume that the surface forms fed to the model at step (1a) are all licit according to the target OT phonotactics the algorithm is being trained on. I assume furthermore that at step (1b), the model assumes an underlying form which is identical to the surface form received at step (1a). This choice makes sense under the assumption that phonological processes are only motivated in order to satisfy phonotactic requirements, so that phonotactically licit forms are those that do not undergo any phonological process, as stated in Tesar's (2008) *surface orientedness* assumption 1.

Assumption 1 (Surface-orientedness) *The underlying OT typology satisfies the three following conditions: **(a)** the set of underlying forms and the set of surface forms coincide; **(b)** all candidate sets contain the faithful candidate; **(c)** a form is licit according to an OT grammar in the typology if and only if it is mapped faithfully into itself by that grammar.* ■

I make no assumptions on how the model chooses the current loser at step (1c) (see Magri 2013f for discussion). Finally, the issue of the choice of the update rule used at step (1e) reduces to the choice of the promotion amount p in the scheme (2). I will come back to this issue at various points in the rest of the paper. The EDRA with the implementation details just specified is called the *EDRA model* of the child's early acquisition of phonotactics.

2. The problem of the restrictiveness of the EDRA model

The EDRA (1) is said to *converge* provided it can only make a finite number of errors, as long as the underlying/winner/loser form triplets $(/x/, [y], \{z\})$ constructed at steps (1a)-(1c) are consistent with some grammar in the underlying OT typology. Furthermore, convergence is said to be *efficient* provided the number of errors grows slowly (polynomially) with the complexity of the underlying OT typology, simply measured in terms of the cardinality of the constraint set. We have a complete characterization of efficient and convergent EDRAs, summarized in the following theorem 1 (Tesar and Smolensky 1998, Boersma 1998, p. 323-327, Boersma 2009, Pater 2008, Magri 2012b,a).

Theorem 1 *An EDRA converges efficiently iff the promotion amount p in (2a) is not too large (namely smaller than ℓ/w, where ℓ is the total number of currently undominated loser-preferring constraints and w is the total number of winner-preferring constraints).* ■

Suppose that the target adult phonology consists of the OT grammar corresponding to a certain ranking $\gg$ of the underlying constraint set. Let L be the corresponding language, namely the set of phonological forms which are licit according to that grammar. Under the assumption that phonotactics is categorical (see for instance Gorman 2013 for recent discussion), the corresponding set of illicit forms is thus the complement of L. In other words, to learn the target adult phonotactics means to learn the language L. Suppose that the EDRA model is trained on this target phonotactics, namely that the surface forms fed to the algorithm at step (1a) all belong to the target language L. As long as the promotion amount is not too large, theorem 1 guarantees that the algorithm will converge to a final ranking vector θ^{fin}. Consider a ranking $\gg^{\text{fin}}$ represented by that final ranking vector, namely that respects the ordering implicitly defined by the relative size of the final ranking values. Let L^{fin} be the corresponding language, namely the set of phonological forms which are licit according to the OT grammar corresponding to the ranking $\gg^{\text{fin}}$. The issue of the computational soundness of the EDRA model of the acquisition of phonotactics thus boils down to the following question: is it possible to provide guarantees that $L^{\text{fin}} = L$, namely that the model succeeds at learning the target phonotactics it is trained on?

Convergence means that the model is *consistent*, namely that the inclusion $L^{\text{fin}} \supseteq L$ holds: there cannot exist any phonological form which is licit according to the target phonotactics (namely, belongs to L) but is incorrectly predicted to be illicit by the phonotactics corresponding to the final ranking vector (namely, does not belong to L^{fin}), as such a form would still be able to trigger an update, contradicting the hypothesis that the algorithm has converged. Thus, the crucial issue concerning the soundness of the EDRA model boils down to the following question: is it possible to provide guarantees that the reverse inclusion $L^{\text{fin}} \subseteq L$ holds as well? In other words, is it possible to provide guarantees that the algorithm will not incorrectly deem as licit forms that are instead illicit according to the target phonotactics? This is the problem of EDRAs' *restrictiveness*.

Theorem 1 provides a powerful theory of EDRAs' convergence and consistency, which crucially holds for any target language in the typology corresponding to any constraint set. The case of restrictiveness turns out to be quite different. In fact, the following theorem 2 (Magri 2013b) says that, for *any* alleged OT learning algorithm (whether error-driven or not), it is possible to construct cases where that algorithm fails at learning the target phonotactics (or else succeeds but is too slow to be efficient). This means in particular that no convergent EDRA can succeed at restrictiveness for any language in the typology corresponding to any constraint set.

Theorem 2 *The problem of the acquisition of phonotactics in OT is intractable: there exists no algorithm (error-driven or not) which is able to solve efficiently an arbitrary instance of the problem.* ■

Such intractability results are usually driven by a small core of tough instances, while the vast majority of the instances of the problem turn out to be quite easy. How could the vast majority of simple instances be separated from the small hard core? Prince and Tesar (2004) and Hayes (2004) conjectured that the tough core of the problem of learning phonotactics has to do with learning the correct relative ranking of the faithfulness constraints. And in Magri (2013a), I show that their conjecture is correct, namely that the problem of learning phonotactics is (easily) solvable (under mild assumptions on the constraint set) if we know the relative ranking of the faithfulness constraints (but of course not the relative ranking of the markedness constraints and the relative ranking between faithfulness and markedness constraints). Yet, we intuitively expect that the relative ranking of the faithfulness constraints mainly determines how illicit forms are repaired. Only in special circumstances does it matter for the distinction between licit and illicit phonological forms. In other words, the relative ranking of the faithfulness constraints is crucial for phonology, but only rarely does it turn out to matter for phonotactics. These considerations suggest that the problem of learning the target phonotactics might actually turn out to be easy in the vast majority of cases, namely all those cases where the relative ranking of the faithfulness constraints does not matter. Easy to the point that even EDRAs will display restrictiveness and thus learn the target adult phonotactics. The rest of this paper formalizes this intuition.

3. $\mathscr{F}$-simple phonotactic patterns

To start with an example, consider the OT typology defined through the set of phonological forms in (3a) and the constraint set in (3b); the *Gen* function only modifies voicing, and is omitted for brevity (Prince and Tesar 2004, Lombardi 1999).

(3) a. $\{$ pa, ba, sa, za, apsa, apza, absa, abza $\}$

$$\text{b.}\quad \left\{ \begin{array}{cc} F_1 = \text{IDENT[STOP-VOICING]} & F_2 = \text{IDENT[FRICATIVE-VOICING]} \\ M_1 = *\text{[STOP-VOICING]} & M_2 = *\text{[FRICATIVE-VOICING]} \\ \multicolumn{2}{c}{M = \text{AGREE[OBSTRUENT-VOICING]}} \end{array} \right\}$$

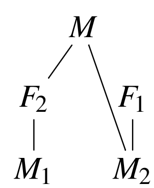
The corresponding OT typology contains in particular the language (4). This language describes the following phonotactics: any form is licit, but for the two forms [apza] and [absa], which display a sequence of obstruents disagreeing in voicing, and thus violate the markedness constraint $M =$ AGREE.

(4) $L = \{$ pa, ba, sa, za apsa, abza $\}$

This language (4) corresponds to two phonologies, depending on how the two illicit forms /apza/ and /absa/ are repaired. One option is (5a): preservation of voicing is more important in stops than fricatives (in the sense that F_1 is ranked above F_2), so that voicing disagreements are repaired by sacrificing fricative-voicing. The reverse option is (5b): preservation of voicing is less important in stops than fricatives (in the sense that F_1 and F_2 are ranked in the reverse order), so that voicing disagreements are repaired by sacrificing stop-voicing.

(5) a. $F_1 \gg F_2 \Longrightarrow \begin{bmatrix} \text{apza} \rightarrow \text{apsa} \\ \text{absa} \rightarrow \text{abza} \end{bmatrix}$

b. $F_2 \gg F_1 \Longrightarrow \begin{bmatrix} \text{apza} \rightarrow \text{abza} \\ \text{absa} \rightarrow \text{apsa} \end{bmatrix}$

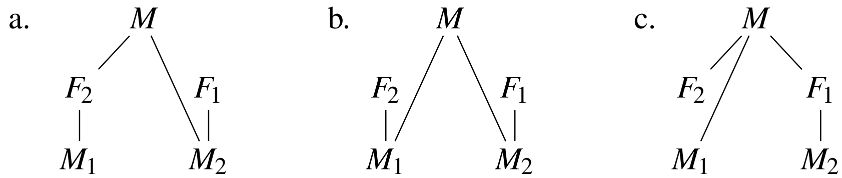
c. 

The relative ranking of the two faithfulness constraints F_1 and F_2 is thus crucial to determine the phonology, namely how the illicit forms /apza/ and /absa/ are repaired.

On the other hand, any ranking that enforces the ranking conditions in (5c) generates the language (4). In other words, any such ranking captures the corresponding phonotactics, namely declares any form licit, but for those ([apza] and [absa]) that display a sequence of obstruents disagreeing in voicing. Crucially, these ranking conditions (5c) say nothing about the relative ranking of the two faithfulness constraints F_1 and F_2: they are consistent with F_1 being ranked above F_2 or vice versa F_1 being ranked below F_2. Thus, although crucial for the phonology, the relative ranking of the faithfulness constraints is irrelevant for the phonotactics in (4). The rest of this section generalizes these considerations.

A ranking is a total order over the constraint set. Thus, a ranking ranks any two faithfulness constraints F_1 and F_2 relative to each other. And there is therefore no way to formalize the intuition that the relative ranking of F_1 and F_2 plays no role. To overcome this problem, let me introduce *partial rankings*, namely rankings that do not necessarily enforce ranking conditions among all constraints. Partial rankings will be denoted by $>, \sqsupset, \ldots$ and variants thereof, to distinguish them from total rankings, which are instead denoted by $\gg, \Supset, \ldots$ and variants thereof. To illustrate, (5c) is an example of a partial ranking over the constraint set in (3b). I repeat it in (6a), together with a couple of additional examples.

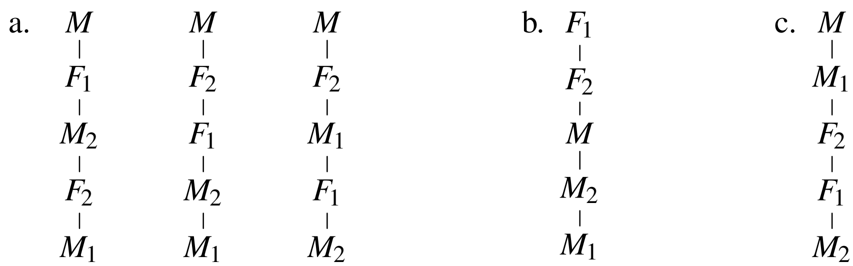
(6)



A partial ranking that does not rank any two faithfulness constraints relative to each other is called $\mathscr{F}$*-simple*.[1] That is for instance the case with the three partial rankings (6) over the constraint set (3b), as none of them imposes any ranking condition between the only two faithfulness constraints F_1 and F_2.

A total ranking $\gg$ is called a *refinement* of a partial ranking $>$ provided the former imposes all the ranking conditions imposed by the latter (plus potentially some more). To illustrate, the total rankings in (7a) are refinements of all the partial rankings in (6), as they respect all the ranking conditions imposed by any of these three partial rankings.

(7)



The total ranking (7b) is a refinement of the partial ranking (6b) but not of the other two partial rankings (6a) and (6c), as the latter two partial rankings require M to be ranked above at least one of the two faithfulness constraints. Finally, the total ranking (7c) is a total refinement of the partial ranking (6c) but not of the other two partial rankings (6a) and (6b), as the latter two partial rankings require F_2 to be ranked above M_1.

Let me say that a partial ranking *generates* a certain language provided each one of its total refinements generates that language, according to the classical definition of OT (see also Yanovich 2012). To illustrate, consider again the language L in (4). Each one of the three total rankings in (7a) generates this language L in the classical sense. As noted above, these three total rankings are refinements of the partial ranking (6a). In effect, any other total refinement of the partial ranking (6a) generates the language L as well. We can thus conclude that this partial ranking (6a) generates the language L. The case of the two other partial rankings (6b) and (6c) is instead different. The partial ranking (6b) does not generate the language L, as it admits the total refinement (7b) which incorrectly predicts forms [apza] and [absa] to be licit as well. And the partial ranking (6c) does not generate the language L either, as it admits the total refinement (7c) which incorrectly predicts the form [abza] to be illicit. Note that the latter two partial rankings (6b) and (6c) admit different total refinements that generate different languages. Thus, these two partial rankings do not generate *any* language, according to the definition assumed here.

A certain phonotactic pattern induces a distinction between licit and illicit phonological forms. As I assume the distinction to be categorical, a phonotactic pattern can be described

[1] Here, I ignore the distinction between *general* and *positional* faithfulness constraints; see Magri (2013e).

through the corresponding set L of licit forms (the set of illicit forms is of course just its complement). The intuition that the relative ranking of the faithfulness constraints does not matter for a certain phonotactic pattern can now be formalized through the condition that the corresponding language L of licit forms is $\mathscr{F}$-simple, namely that it can be generated by a partial ranking which is itself $\mathscr{F}$-simple, namely a partial ranking that does not rank any two faithfulness constraints relative to each other. To illustrate, the language L in (4) describes an $\mathscr{F}$-simple phonotactic pattern.

4. Restrictiveness of the EDRA model on $\mathscr{F}$-simple phonotactic patterns

I am now in a position to state the question addressed in this paper. Suppose that the EDRA model is trained on a target language L which happens to be $\mathscr{F}$-simple. In this case, can we guarantee that the model is restrictive and thus succeeds at learning exactly the target language at convergence? In this section, I provide a positive answer to this question, under two extremely mild assumptions on the underlying OT typology (besides Tesar's 2008 surface-orientedness assumption 1 stated above).

Suppose that the set of faithfulness constraints is impoverished — to take an extreme case, suppose it is empty. Then of course the issue of the proper relative ranking of the faithfulness constraints does not arise at all. In this case, the assumption that a certain phonotactic pattern is $\mathscr{F}$-simple thus has no bite. These considerations suggest that we need in place an assumption which ensures that the set of faithfulness constraints is rich enough. This is the purpose of the following assumption 2.

Assumption 2 ($\mathscr{F}$-discernibility) *For any underlying form $/x/$ and any non faithful candidate $[y]$ of $/x/$, there exists a faithfulness constraint F such that $F(/x/,[y]) \neq 0$.* ■

Suppose that a phonological form is a candidate of another phonological form. This means that the candidate can be obtained from the underlying form by performing certain phonological operations, such as changing certain feature values (e.g., devoicing), inserting an epenthetic segment (e.g., breaking a consonant cluster through an epenthetic vowel), deleting a segment (e.g., simplifying a consonant cluster into a singleton), etcetera. Plausibly, these operations can be inverted: feature values can be changed back to their original setting, epenthetic segments can be deleted, deleted segments can be epenthesized back, and so on. By applying these reverse operations, the underlying form can be obtained back from its candidate. These considerations lead to the assumption 3 that the candidacy relation *Gen* is symmetric.[2]

Assumption 3 (Symmetry) *The candidacy relation Gen is symmetric: $[y]$ is a candidate of $/x/$ if and only if $[x]$ is vice versa a candidate of $/y/$.* ■

Suppose now that the EDRA model described in section 1 is trained on a target language L which is $\mathscr{F}$-simple and belongs to an OT typology which satisfies the three as-

[2]In general, the generating function *Gen* is a relation between the two possibly different sets of underlying and surface forms. In this general setting, assumption 3 that *Gen* be symmetric makes no sense. This assumption only makes sense based on condition (a) of assumption 1, which requires the two sets of underlying and surface forms to coincide.

sumptions 1, 2, and 3. The markedness constraints start out high, with an initial ranking value equal to a constant θ^{init}, usually much larger than the number m of markedness constraints. The faithfulness constraints instead start out low, with a null initial ranking value. Throughout learning, the faithfulness constraints will raise above zero, in case the algorithm adopts a non-null promotion amount $p > 0$. Theorem 3 ensures that the EDRA manages to learn the target phonotactics, as long as the faithfulness constraints don't raise too high, namely their ranking values remain smaller by at least m than the initial ranking value θ^{init} of the markedness constraints, as stated in the crucial condition (8). A sketch of the proof of theorem 3 is provided in section 5; see Magri (2013e) for full details.

Theorem 3 *Assume that the underlying OT typology is surface-oriented, in the sense of assumption 1; that the set of faithfulness constraints is rich enough to single out non-faithful candidates, in the sense of assumption 2; and that the candidacy relation is symmetric, as required by assumption 3. Consider a language L in the typology which is $\mathscr{F}$-simple, namely a language generated by a partial ranking that does not rank any two faithfulness constraints relative to each other. Let a convergent EDRA model run on this training language L until it converges. Assume that the ranking value θ_F of any faithfulness constraint F at any time in the run satisfies the crucial condition (8), where m is the total number of markedness constraints and θ^{init} is their initial ranking value.*

(8) $\theta_F \leq \theta^{init} - m$

Then, the language L^{fin} generated by (an arbitrary refinement of) the final ranking vector entertained by the EDRA at convergence coincides with the target language L the EDRA has been trained on. ■

Plausibly, the assumption that the language L of licit forms is $\mathscr{F}$-simple turns out to be satisfied for the vast majority of languages in a given typology. Thus, theorem 3 guarantees that the EDRA model succeeds at learning the target phonotactics in the vast majority of cases within a given typology. Furthermore, the two assumptions 2, and 3 required by the theorem are extremely mild. Thus, theorem 3 provides guarantees of succeeds on $\mathscr{F}$-simple languages under virtually no restrictions on the underlying constraint set. The main assumption of the theorem is the condition (8) that the faithfulness constraints don't raise too high. This assumption is trivially satisfied when the promotion amount p in (2a) is set equal to zero: in this case, the faithfulness constraints are not promoted at all, and thus cannot be promoted too high. Magri (2013c) discusses this condition (8) in full detail, showing that it is not incompatible with a non-null promotion amount $p > 0$.

5. Sketch of the proof of theorem 3

Suppose that the EDRA model is trained on a target language L which is $\mathscr{F}$-simple, namely is generated by a partial ranking $>$ that does not rank any two faithfulness constraints relative to each other. It is immediate to see that $>$ admits a total refinement $\gg$ with the *tripartite* shape (9), for some partition of the set $\mathscr{M}$ of markedness constraints into two disjoint sets $\mathscr{M}_{\text{top}}$ and $\mathscr{M}_{\text{bottom}}$.

(9) $\mathscr{M}_{\text{top}} \gg \mathscr{F} \gg \mathscr{M}_{\text{bottom}}$

In other words, the total ranking $\gg$ assigns the markedness constraints in $\mathscr{M}_{\text{top}}$ at the top of the ranking, with a certain relative ranking among them; then come the faithfulness constraints, with a certain relative ranking among them; and finally come the remaining markedness constraints $\mathscr{M}_{\text{bottom}}$, again with a certain relative ranking among them.

Suppose that the EDRA model is run on this training language until it converges, namely until it can make no more mistakes. Consider an arbitrary ranking $\gg^{\text{fin}}$ represented by the final ranking vector entertained by the model at convergence. I want to prove that $L^{\text{fin}} = L$, namely that the language L^{fin} generated by this final total ranking $\gg^{\text{fin}}$ coincides with the target language L the model has been trained on. As noted in section 2, *convergence* ensures that the algorithm will learn to recognize licit forms as such. In other words, convergence ensures the inclusion $L \subseteq L^{\text{fin}}$. I thus need to prove the reverse inclusion $L \supseteq L^{\text{fin}}$: if a certain form x does not belong to the target language L, then it does not belong to the final language L^{fin} either, thus ensuring *restrictiveness*.

Since the phonological form x does not belong to the target language L and since the total ranking $\gg$ generates the language L, then the OT grammar corresponding to the ranking $\gg$ cannot let x surface faithfully but rather needs to neutralize it to some different phonological form y that belongs instead to the target language L and is therefore licit according to the target phonotactics. This means in particular that y is a better candidate for x than x itself according to this ranking $\gg$. In other words, this ranking $\gg$ is consistent with the underlying/winner/loser form triplet $(/x/, [y], [\not{x}])$. Consider the constraints that are loser-preferring relative to this triplet. Without loss of generality, assume that the loser-preferring faithfulness constraints are $F_1, \ldots, F_h$ and the loser-preferring markedness constraints are $M_1, \ldots, M_k$. As the tripartite ranking $\gg$ is consistent with this underlying/winner/loser form triplet $(/x/, [y], [\not{x}])$, then there has got to exist some constraint that is winner-preferring relative to that triplet and furthermore is $\gg$-ranked above all these loser-preferrers $F_1, \ldots, F_h$ and $M_1, \ldots, M_k$. And of course this winner-preferring constraint must be a markedness constraint (call it M), as no faithfulness constraint can ever be winner-preferring relative to a triplet such as $(/x/, [y], [\not{x}])$, whose loser is faithful to the underlying form. In conclusion, the tripartite ranking $\gg$ satisfies the ranking conditions in (10).

(10)
M — a markedness constraint W-preferring for $(/x/, [y], [\not{x}])$
(b) $M_1, \ldots, M_k$ — all markedness constraints L-preferring for $(/x/, [y], [\not{x}])$
(a) $F_1, \ldots, F_h$ — all faithfulness constraints L-preferring for $(/x/, [y], [\not{x}])$

I now work separately on the two target ranking conditions (10a) and (10b).

By assumption 2, there has got to exist at least one faithfulness constraint that is violated by the mapping of the underlying form x into the non faithful candidate y. Of course, that faithfulness constraint is loser-preferring relative to the underlying/winner/loser form triplet $(/x/, [y], [\not{x}])$. Since the markedness constraint M is $\gg$-ranked above that faithfulness constraint by (10a) and since this ranking $\gg$ has the special tripartite shape $\mathscr{M}_{\text{top}} \gg \mathscr{F} \gg \mathscr{M}_{\text{bottom}}$ in (9), then this markedness constraint M must belong to the set $\mathscr{M}_{\text{top}}$ of markedness constraints that are $\gg$-ranked above every faithfulness constraint.

All markedness constraints start high with an initial large ranking value θ^{init}, both those in $\mathscr{M}_{\text{top}}$ and those in $\mathscr{M}_{\text{bottom}}$. The markedness constraints in $\mathscr{M}_{\text{bottom}}$ might drop a long way in order to end up underneath the faithfulness constraints. The case of the markedness constraints in $\mathscr{M}_{\text{top}}$ is instead quite different. Indeed, an argument due to Tesar and Smolensky (1998) shows that an EDRA (which only demotes loser-preferring constraints which are undominated) cannot demote the markedness constraints in $\mathscr{M}_{\text{top}}$ too much, namely their ranking value can never drop below the crucial threshold $\theta^{\text{init}} - m$, where m is the total number of markedness constraints. On the other hand, the faithfulness constraints start low, namely with a null initial ranking value. Suppose the EDRA manages not to promote them too high, in the sense that their ranking values never pass the forbidden threshold of $\theta^{\text{init}} - m$, as required by the crucial condition (8). Then, the EDRA will manage to keep the markedness constraints in $\mathscr{M}_{\text{top}}$ higher up (above the crucial threshold) and the faithfulness constraints lower down (below the crucial threshold), thus enforcing the desired ranking conditions $\mathscr{M}_{\text{top}} \gg \mathscr{F}$. I can thus conclude that any refinement $\gg^{\text{fin}}$ of the final ranking vector entertained by the EDRA indeed ranks the markedness constraint M above any faithfulness constraint, thus enforcing the ranking condition (10a).

Let me turn to the ranking condition (10b) that the constraint M be ranked above all markedness constraints $M_1, M_2, \ldots, M_k$ that are loser-preferring relative to the triplet $(/x/, [y], [\not{x}])$. Suppose $\gg^{\text{fin}}$ fails to enforce these ranking conditions. For concreteness, let's say that $\gg^{\text{fin}}$ fails to rank underneath M the loser-preferring markedness constraint M_1. As $\gg^{\text{fin}}$ is a total ranking, this means in turn that M_1 must be ranked above M, as represented in (11), where I use a curly line ("$\sim\!\sim$") to represent the ranking $\gg^{\text{fin}}$, while I use a straight line as in (10) to represent the ranking $\gg$.

(11)

M_1	— L-preferring for $(/x/, [y], [\not{x}])$	— L-preferring for $(/y/, [y], [\not{x}])$
M	— W-preferring for $(/x/, [y], [\not{x}])$	
$M_2, \ldots, M_k$	— L-preferring for $(/x/, [y], [\not{x}])$	
$\mathscr{F}$	— all $\mathscr{F}$ constraints	

As y is a candidate of x, then x is vice versa a candidate of y, by the assumption 3 that the candidacy relation is symmetric. As M_1 is a markedness constraint that is loser-preferring relative to the triplet $(/x/, [y], [\not{x}])$, then it is also loser-preferring relative to the triplet $(/y/, [y], [\not{x}])$, as the two triplets only differ for the underlying form (and markedness constraints do not look at the underlying forms). I have indicated this in (11), where the rightmost column classifies constraints relative to the triplet $(/y/, [y], [\not{x}])$

By convergence, the final ranking $\gg^{\text{fin}}$ is consistent with the target language L. As y belongs to this language L by hypothesis, then the final ranking $\gg^{\text{fin}}$ is consistent with y. By the surface-orientedness assumption 1, this means in turn that the final ranking $\gg^{\text{fin}}$ is consistent with the underlying/winner/loser form triplet $(/y/, [y], [\not{\ldots}])$, no matter how we choose a non-faithful loser candidate for the underlying form y. As x is a non-faithful candidate of y, the final ranking $\gg^{\text{fin}}$ is thus in particular consistent with the underlying/winner/loser form triplet $(/y/, [y], [\not{x}])$. As M_1 is loser-preferring relative to this triplet, then the final ranking $\gg^{\text{fin}}$ needs to rank above it some constraint C that is winner-preferring relative to this triplet $(/y/, [y], [\not{x}])$, as indicated in (12).

(12)

C	—W-preferring for $(/x/, [y], [\not{x}])$	W-preferring for $(/y/, [y], [\not{x}])$
M_1	— L-preferring for $(/x/, [y], [\not{x}])$	L-preferring for $(/y/, [y], [\not{x}])$
M	— W-preferring for $(/x/, [y], [\not{x}])$	
$M_2, \dots, M_k$	— L-preferring for $(/x/, [y], [\not{x}])$	
$\mathscr{F}$	— all $\mathscr{F}$ constraints	

As the final ranking $\gg^{\text{fin}}$ ranks all the faithfulness constraints underneath M_1, then this constraint C ranked above M_1 cannot be a faithfulness constraint and must instead be a markedness constraint. As C is a markedness constraint, the fact that it is winner-preferring relative to the underlying/winner/loser form triplet $(/y/, [y], [\not{x}])$ guarantees that it is also winner-preferring relative to the original triplet $(/x/, [y], [\not{x}])$, as the two triplets only differ for the underlying form.[3] In conclusion, the final ranking $\gg^{\text{fin}}$ satisfies the ranking conditions (12). These are analogous to the ranking condition (10) satisfied by the target ranking $\gg$, only with the markedness constraint C in place of M. I thus conclude that the final ranking $\gg^{\text{fin}}$ as well deems the form x illicit, just as the target ranking.

6. Conclusions

This paper has looked at guarantees that the OT error-driven model succeeds at learning the target adult phonotactics, when trained on licit forms only but no alternations. I have focused on the special case where the target phonotactics does not require any specific relative ranking of the faithfulness constraints, formalized in section 3. The main result of this paper is theorem 3, which ensures that in this case the model succeeds under very mild assumptions on the constraint set, as long as the faithfulness constraints don't raise too high. Magri (2013e) contains a more detailed presentation of this result.

The crucial assumption that this result relies on is that the faithfulness constraints don't raise too high. Magri (2013c) discusses this assumption in detail. It shows that the height reached by the faithfulness constraints can be controlled also for implementations of the OT error-driven model that perform constraint promotion as well as demotion (if no constraint promotion is performed, the faithfulness constraint don't raise at all, and thus cannot raise too high). Taken together, these two papers thus guarantee restrictiveness of the OT error-driven model when the relative ranking of the faithfulness constraints does not matter. And these guarantees require virtually no assumptions on the constraint set.

What about the complementary case, namely those phonotactic patterns which instead do require a specific relative ranking of the faithfulness constraints? The intractability result recalled above as theorem 2 says in particular that guarantees for restrictiveness in this complementary case cannot be provided without making proper assumptions on the

[3]The fact that this constraint C is a markedness constraint is crucial here. In fact, this step of the reasoning would not hold in the case of a non-symmetric faithfulness constraints, such as DEP or MAX.

constraint set. This result thus motivates the following research strategy: focus on specific families of constraint sets and exploit the special properties of the typologies they induce in order to provide guarantees for restrictiveness of the OT error-driven model. Magri (2013d) provides an initial implementation of this research strategy for a large family of OT constraints for inventory segmental phonotactics. Building on the latter results, Magri (2013a) shows that, despite its very limited computational resources, the OT error-driven model of the acquisition of phonotactics substantially outperforms recent batch models (Prince and Tesar 2004, Hayes 2004).

References

Boersma, Paul. 1997. How we learn variation, optionality and probability. In *Proceedings of the Institute of Phonetic Sciences (IFA) 21*, ed. Rob van Son, 43–58. University of Amsterdam: Institute of Phonetic Sciences.

Boersma, Paul. 1998. Functional phonology. Doctoral Dissertation, University of Amsterdam, The Netherlands. The Hague: Holland Academic Graphics.

Boersma, Paul. 2009. Some correct error-driven versions of the constraint demotion algorithm. *Linguistic Inquiry* 40:667–686.

Fenson, Larry, Philip S. Dale, J. Steven Reznick, Elizabeth Bates, Donna J. Thal, and Stephan J. Pethick. 1994. *Variability in early communicative development*. Monographs of the Society for Research in Child Development 59 (Serial No. 242, Vol. 59, No 5).

Gorman, Kyle. 2013. Generative phonotactics. Doctoral Dissertation, University of Pennsylvania.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: The early stages. In *Constraints in phonological acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld, 158–203. Cambridge: Cambridge University Press.

Jusczyk, P. W., A. D. Friederici, J. M. I. Wessels, V. Y. Svenkerud, and A. Jusczyk. 1993. Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language* 32:402–420.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17:267–302.

Magri, Giorgio. 2012a. Constraint promotion: not only convergent, but also efficient. In *Proceedings of the 48th annual conference of the Chicago Linguistics Society*.

Magri, Giorgio. 2012b. Convergence of error-driven ranking algorithms. *Phonology* 29:213–269.

Magri, Giorgio. 2013a. Batch versus error-driven models of the acquisition of phonotactics: David defeats Goliath. Manuscript, CNRS and University of Paris 8.

Magri, Giorgio. 2013b. The complexity of learning in OT and its implications for the acquisition of phonotactics. *Linguistic Inquiry* 44.3:433–468.

Magri, Giorgio. 2013c. The error-driven ranking model of the acquisition of phonotactics: controlling the height of the faithfulness constraints. Manuscript, CNRS and University of Paris 8.

Magri, Giorgio. 2013d. The error-driven ranking model of the acquisition of phonotactics: restrictiveness in the case of inventory segmental phonotactics. Manuscript, CNRS and University of Paris 8.

Magri, Giorgio. 2013e. The error-driven ranking model of the acquisition of phonotactics: restrictiveness when the relative ranking of the faithfulness constraints does not matter. Manuscript, CNRS and University of Paris 8.

Magri, Giorgio. 2013f. A note on the GLA's choice of the current loser from the perspective of factorizability. *Journal of Logic, Language, and Information* 22:231–247.

Pater, Joe. 2008. Gradual learning and convergence. *Linguistic Inquiry* 39.2:334–345.

Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Blackwell. As Technical Report CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder, and Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ, April 1993. Also available as ROA 537 version.

Prince, Alan, and Bruce Tesar. 2004. Learning phonotactic distributions. In *Constraints in phonological acquisition*, ed. R. Kager, J. Pater, and W. Zonneveld, 245–291. Cambridge University Press.

Smolensky, Paul. 1996. The initial state and Richness of the Base in Optimality Theory. John Hopkins Technical Report.

Tesar, Bruce. 2008. Output-driven maps. Ms., Rutgers University. Also available as ROA-956.

Tesar, Bruce, and Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29:229–268.

Yanovich, Igor. 2012. The logic of ot rankings. MIT manuscript.

SFL UMR 7023 "Structures formelles du langage"
CNRS, University of Paris 8
Paris, France

magrigrg@gmail.com