The error-driven ranking model of the acquisition of phonotactics: some computational results

Giorgio Magri

SFL UMR 7023 (CNRS and University of Paris 8)

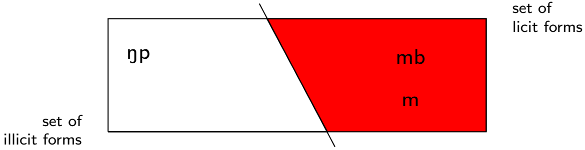
Département de Linguistique
Université de Genève
8 March 2016

1.

Error-driven learning of phonotactics

Phonology = phonotactics + repairs

- Phonotactics: knowledge of the language specific acceptability of forms, construed as a categorical distinction: licit vs illicit [Gorman 2013]

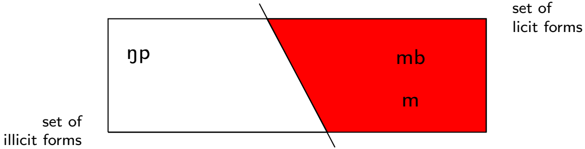


- Repairs: knowledge of the language specific repairs of illicit forms, as revealed by alternations from URs to SRs: [Kenstowicz and Kisseberth 1977]

Indonesian:	*Quechua:*
/məŋ+pilih/ → [məmilih] 'to vote'	/kaŋ+pa/ → [kamba] 'yours'
/məŋ+tulis/ → [mənulis] 'to write'	/wakiŋ+ta/ → [wakinda] 'the others'

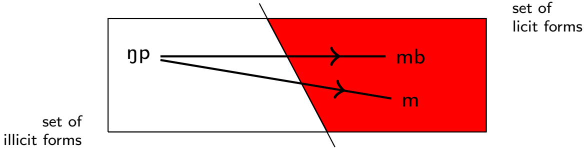
Phonology = phonotactics + repairs

- Phonotactics: knowledge of the language specific acceptability of forms, construed as a categorical distinction: licit vs illicit [Gorman 2013]



- Repairs: knowledge of the language specific repairs of illicit forms, as revealed by alternations from URs to SRs: [Kenstowicz and Kisseberth 1977]

Indonesian:	*Quechua:*
/mə**ŋ**+**p**ilih/ → [mə**m**ilih] 'to vote'	/ka**ŋ**+**p**a/ → [ka**mb**a] 'yours'
/mə**ŋ**+**t**ulis/ → [mə**n**ulis] 'to write'	/waki**ŋ**+**t**a/ → [waki**nd**a] 'the others'



Learnability logics and acquisition phenomenology

- **Knowledge of repairs ⇒ knowledge of phonotactics:**
  - if you know alternations, you know what is licit and what is not
  - i.e. phonotactics is the *range* of a phonological grammar G

- **Learnability logics:** [Prince and Tesar 2008]
  - start by focusing on a smaller aspect of the problem of learning phonology, namely start with the problem of learning the phonotactics
  - then use knowledge of phonotactics to bootstrap into knowledge of the whole system of phonological alternations

- **Acquisition phenomenology:**
  - nine-month-olds react differently to illicit sounds [Jusczyk et al. 1993]
  - they thus display knowledge of the target phonotactics
  - at a stage when morphology is lagging behind [Kazazis 1969; Hayes 2004]
  - the child has still no access to phonological alternations

Learnability logics and acquisition phenomenology

- **Knowledge of repairs ⇒ knowledge of phonotactics:**
  - if you know alternations, you know what is licit and what is not
  - i.e. phonotactics is the *range* of a phonological grammar G

- **Learnability logics:** [Prince and Tesar 2008]
  - start by focusing on a smaller aspect of the problem of learning phonology, namely start with the problem of learning the phonotactics
  - then use knowledge of phonotactics to bootstrap into knowledge of the whole system of phonological alternations

- **Acquisition phenomenology:**
  - nine-month-olds react differently to illicit sounds [Jusczyk et al. 1993]
  - they thus display knowledge of the target phonotactics
  - at a stage when morphology is lagging behind [Kazazis 1969; Hayes 2004]
  - the child has still no access to phonological alternations

Learnability logics and acquisition phenomenology

- **Knowledge of repairs ⇒ knowledge of phonotactics:**
  - if you know alternations, you know what is licit and what is not
  - i.e. phonotactics is the *range* of a phonological grammar G

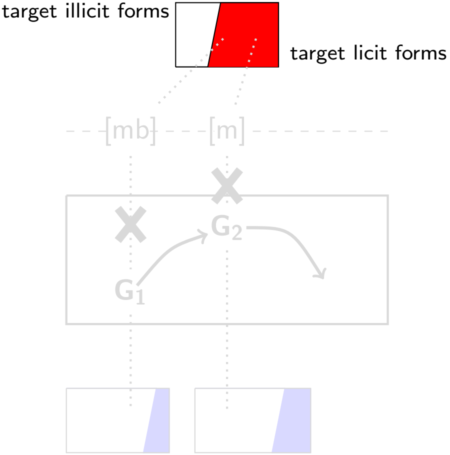
- **Learnability logics:** [Prince and Tesar 2008]
  - start by focusing on a smaller aspect of the problem of learning phonology, namely start with the problem of learning the phonotactics
  - then use knowledge of phonotactics to bootstrap into knowledge of the whole system of phonological alternations

- **Acquisition phenomenology:**
  - nine-month-olds react differently to illicit sounds [Jusczyk et al. 1993]
  - they thus display knowledge of the target phonotactics
  - at a stage when morphology is lagging behind [Kazazis 1969; Hayes 2004]
  - the child has still no access to phonological alternations

Learnability logics and acquisition phenomenology

- **Knowledge of repairs ⇒ knowledge of phonotactics:**
  - if you know alternations, you know what is licit and what is not
  - i.e. phonotactics is the *range* of a phonological grammar G
- **Learnability logics:** [Prince and Tesar 2008]
  - start by focusing on a smaller aspect of the problem of learning phonology, namely start with the problem of learning the phonotactics
  - then use knowledge of phonotactics to bootstrap into knowledge of the whole system of phonological alternations
- **Acquisition phenomenology:**
  - nine-month-olds react differently to illicit sounds [Jusczyk et al. 1993]
  - they thus display knowledge of the target phonotactics
  - at a stage when morphology is lagging behind [Kazazis 1969; Hayes 2004]
  - the child has still no access to phonological alternations

Error-driven learning



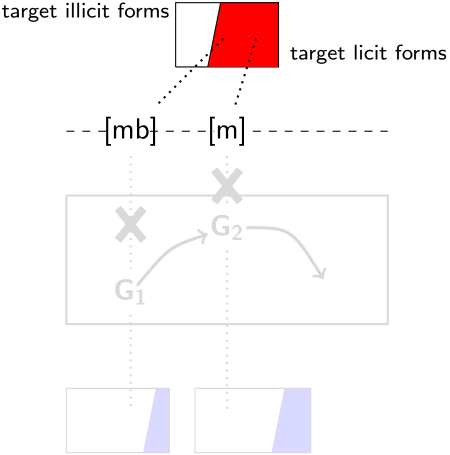
target adult phonotactics

stream of input data

space of all possible grammars

currently predicted phonotactics

Error-driven learning



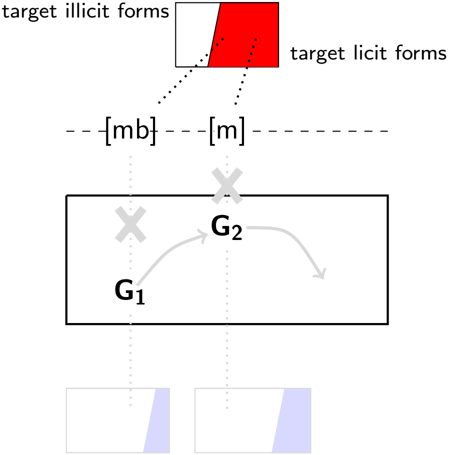
target adult phonotactics

stream of input data

space of all possible grammars

currently predicted phonotactics

Error-driven learning



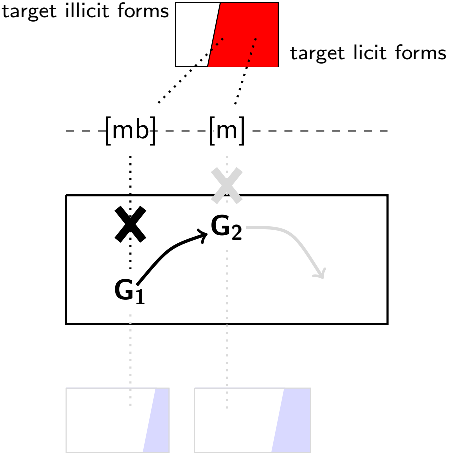
target adult phonotactics

stream of input data

space of all possible grammars

currently predicted phonotactics

Error-driven learning



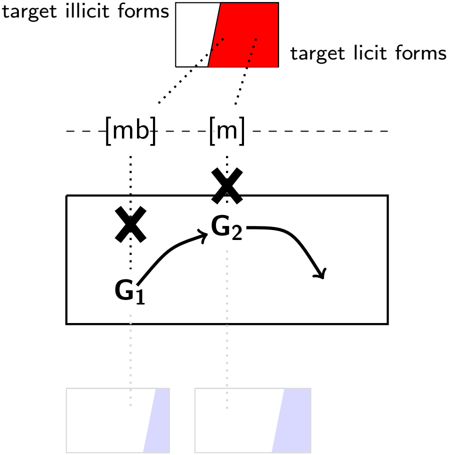
target adult phonotactics

stream of input data

space of all possible grammars

currently predicted phonotactics

Error-driven learning



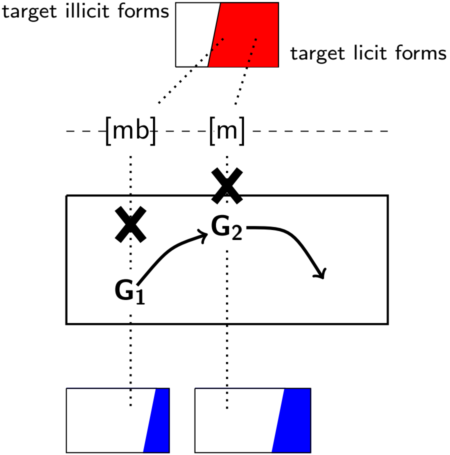
target adult phonotactics

stream of input data

space of all possible grammars

currently predicted phonotactics

Error-driven learning



target adult phonotactics

stream of input data

space of all possible grammars

currently predicted phonotactics

Two modeling virtues

- **Memory-free**: [Gibson and Wexler 1994]
  - the error-driven learner doesn't keep track of previously seen data
  - doesn't need a lexicon of stored forms
  - it is therefore suitable to model early acquisition stages

- **Gradual**: [McLeod et al. 2001]
  - the error-driven learner describes a sequence of grammars
  - which can be matched with child acquisition paths
  - it is therefore suitable to model acquisition gradualness

Two modeling virtues

- **Memory-free**: [Gibson and Wexler 1994]
  - the error-driven learner doesn't keep track of previously seen data
  - doesn't need a lexicon of stored forms
  - it is therefore suitable to model early acquisition stages
- **Gradual**: [McLeod et al. 2001]
  - the error-driven learner describes a sequence of grammars
  - which can be matched with child acquisition paths
  - it is therefore suitable to model acquisition gradualness

Goal of the talk

- **Today**, I will focus [Cesa-Bianchi and Lugosi 2006; Tesar and Smolensky 1998; Boersma 1998]
  - on the early purely phonotactic learning stage
  - through error-driven learning
  - within the OT implementation of constraint-based phonology

⇒ error-driven ranking algorithms (EDRAs) for phonotactic learning

- **Does it work?** Are there guarantees that the learned phonotactics coincides with the target phonotactics?

- **The right initial question:** "a [learning model] that is powerful enough to account for the *fact* of language acquisition may be a more promising first approximation of an ultimately viable theory than one that is able to describe the *course* of language acquisition, which has been the traditional focus of developmental psycholinguistics" [Pinker 1979]

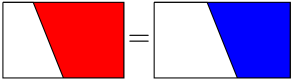
Goal of the talk

- **Today**, I will focus [Cesa-Bianchi and Lugosi 2006; Tesar and Smolensky 1998; Boersma 1998]
  - on the early purely phonotactic learning stage
  - through error-driven learning
  - within the OT implementation of constraint-based phonology

  ⇒ error-driven ranking algorithms (EDRAs) for phonotactic learning

- **Does it work?** Are there guarantees that the learned phonotactics coincides with the target phonotactics?

target phonotactics



learned phonotactics

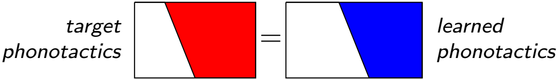
- **The right initial question:** "a [learning model] that is powerful enough to account for the *fact* of language acquisition may be a more promising first approximation of an ultimately viable theory than one that is able to describe the *course* of language acquisition, which has been the traditional focus of developmental psycholinguistics" [Pinker 1979]

Goal of the talk

- **Today**, I will focus [Cesa-Bianchi and Lugosi 2006; Tesar and Smolensky 1998; Boersma 1998]
  - on the early purely phonotactic learning stage
  - through error-driven learning
  - within the OT implementation of constraint-based phonology

  ⇒ error-driven ranking algorithms (EDRAs) for phonotactic learning

- **Does it work?** Are there guarantees that the learned phonotactics coincides with the target phonotactics?



- **The right initial question:** "a [learning model] that is powerful enough to account for the *fact* of language acquisition may be a more promising first approximation of an ultimately viable theory than one that is able to describe the *course* of language acquisition, which has been the traditional focus of developmental psycholinguistics" [Pinker 1979]

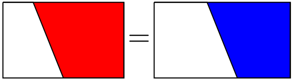
Goal of the talk

- **Today**, I will focus [Cesa-Bianchi and Lugosi 2006; Tesar and Smolensky 1998; Boersma 1998]
  - on the early purely phonotactic learning stage
  - through error-driven learning
  - within the OT implementation of constraint-based phonology

$\Rightarrow$ error-driven ranking algorithms (EDRAs) for phonotactic learning

- **Does it work?** Are there guarantees that the learned phonotactics coincides with the target phonotactics?

target phonotactics



learned phonotactics

- **Two sub-questions:**
  - Consistency:

target

learned

  - Restrictiveness:

target

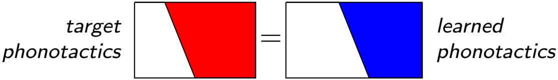
learned

Goal of the talk

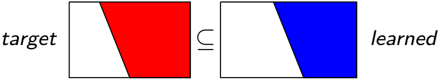
- **Today**, I will focus [Cesa-Bianchi and Lugosi 2006; Tesar and Smolensky 1998; Boersma 1998]
  - on the early purely phonotactic learning stage
  - through error-driven learning
  - within the OT implementation of constraint-based phonology

  ⇒ error-driven ranking algorithms (EDRAs) for phonotactic learning

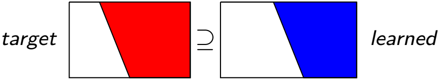
- **Does it work?** Are there guarantees that the learned phonotactics coincides with the target phonotactics?



- **Two sub-questions:**
  - Consistency:



  - Restrictiveness:



2.
OT implementation

Shortest overview of constraint-based phonology

- The core data unit is a comparison between two candidate mappings for the same underlying form: (/ŋp/, [m]) versus (/ŋp/, [mb])
- Languages differ in which candidate beats the other (stricken out):
  - (/ŋp/, [m]) beats (/ŋp/, ~~[mb]~~) according to Indonesian
  - (/ŋp/, [mb]) beats (/ŋp/, ~~[m]~~) according to Quechua
- A set of constraints map these phonological comparisons into $\mathbb{R}^n$

(/ŋp/, [m]) beats (/ŋp/, ~~[mb]~~) $\Longrightarrow$

Shortest overview of constraint-based phonology

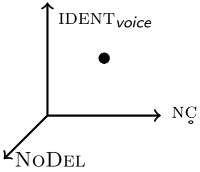
- The core data unit is a comparison between two candidate mappings for the same underlying form: (/ŋp/, [m]) versus (/ŋp/, [mb])
- Languages differ in which candidate beats the other (stricken out):
  - (/ŋp/, [m]) beats (/ŋp/, ~~[mb]~~) according to Indonesian
  - (/ŋp/, [mb]) beats (/ŋp/, ~~[m]~~) according to Quechua
- A set of constraints map these phonological comparisons into $\mathbb{R}^n$

(/ŋp/, [m]) beats (/ŋp/, ~~[mb]~~) $\Longrightarrow$

Shortest overview of constraint-based phonology

- The core data unit is a comparison between two candidate mappings for the same underlying form: (/ŋp/, [m]) versus (/ŋp/, [mb])
- Languages differ in which candidate beats the other (stricken out):
  - (/ŋp/, [m]) beats (/ŋp/, ~~[mb]~~) according to Indonesian
  - (/ŋp/, [mb]) beats (/ŋp/, ~~[m]~~) according to Quechua
- A set of constraints map these phonological comparisons into $\mathbb{R}^n$

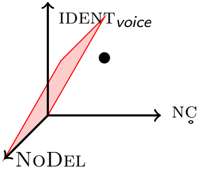
(/ŋp/, [m]) beats (/ŋp/, ~~[mb]~~) $\Longrightarrow$



Shortest overview of constraint-based phonology

- The core data unit is a comparison between two candidate mappings for the same underlying form: (/ŋp/, [m]) versus (/ŋp/, [mb])
- Languages differ in which candidate beats the other (stricken out):
  - (/ŋp/, [m]) beats (/ŋp/, ~~[mb]~~) according to Indonesian
  - (/ŋp/, [mb]) beats (/ŋp/, ~~[m]~~) according to Quechua
- A set of constraints map these phonological comparisons into $\mathbb{R}^n$

(/ŋp/, [m]) beats (/ŋp/, ~~[mb]~~) $\Longrightarrow$



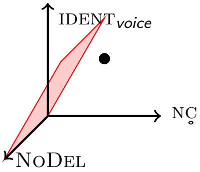
- A grammar corresponds to a hyperplane that leaves all its comparisons on one side
- A grammar is described by assigning constraint weights which define the corresponding hyperplane

[Prince and Smolensky 2004]

Shortest overview of constraint-based phonology

- The core data unit is a comparison between two candidate mappings for the same underlying form: (/ŋp/, [m]) versus (/ŋp/, [mb])
- Languages differ in which candidate beats the other (stricken out):
  - (/ŋp/, [m]) beats (/ŋp/, ~~[mb]~~) according to Indonesian
  - (/ŋp/, [mb]) beats (/ŋp/, ~~[m]~~) according to Quechua
- A set of constraints map these phonological comparisons into $\mathbb{R}^n$

(/ŋp/, [m]) beats (/ŋp/, ~~[mb]~~) $\Longrightarrow$



- A grammar corresponds to a hyperplane that leaves all its comparisons on one side
- A grammar is described by assigning constraint weights which define the corresponding hyperplane

[Prince and Smolensky 2004]

Typological restrictions on the hyperplanes

- Are there restrictions on the hyperplanes that correspond to actual phonological grammars?

- **The OT empirical generalization:** [Prince and Smolensky 2004]
  natural language phonologies display no additive/gang-up effects

- Equivalently, natural language phonologies...
  - ...correspond to hyperplanes whose weights decay exponentially
  - ...correspond to very titled hyperplanes
  - ...can be re-parametrized with rankings
  - ...enforce strict domination

  This is a currently open typological debate [Pater 2007]

- "That strict domination governs grammatical constraint interaction is not currently explained" [Prince and Smolensky 1997]

Typological restrictions on the hyperplanes

- Are there restrictions on the hyperplanes that correspond to actual phonological grammars?

- **The OT empirical generalization:** [Prince and Smolensky 2004]
  natural language phonologies display no additive/gang-up effects

- Equivalently, natural language phonologies. . .
  - . . . correspond to hyperplanes whose weights decay exponentially
  - . . . correspond to very titled hyperplanes
  - . . . can be re-parametrized with rankings
  - . . . enforce strict domination

  This is a currently open typological debate [Pater 2007]

- "That strict domination governs grammatical constraint interaction is not currently explained" [Prince and Smolensky 1997]

Typological restrictions on the hyperplanes

- Are there restrictions on the hyperplanes that correspond to actual phonological grammars?

- **The OT empirical generalization:** [Prince and Smolensky 2004]
  natural language phonologies display no additive/gang-up effects

- Equivalently, natural language phonologies...
  - ...correspond to hyperplanes whose weights decay exponentially
  - ...correspond to very titled hyperplanes
  - ...can be re-parametrized with rankings
  - ...enforce strict domination

  This is a currently open typological debate [Pater 2007]

- "That strict domination governs grammatical constraint interaction is not currently explained" [Prince and Smolensky 1997]

Typological restrictions on the hyperplanes

- Are there restrictions on the hyperplanes that correspond to actual phonological grammars?

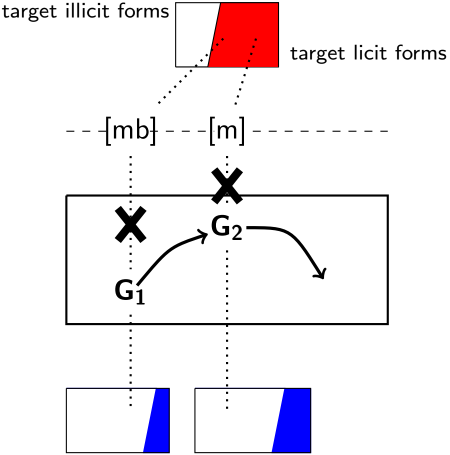
- **The OT empirical generalization:** [Prince and Smolensky 2004]
  natural language phonologies display no additive/gang-up effects

- Equivalently, natural language phonologies...
  - ...correspond to hyperplanes whose weights decay exponentially
  - ...correspond to very titled hyperplanes
  - ...can be re-parametrized with rankings
  - ...enforce strict domination

  This is a currently open typological debate [Pater 2007]

- "That strict domination governs grammatical constraint interaction is not currently explained" [Prince and Smolensky 1997]

Error-driven learning



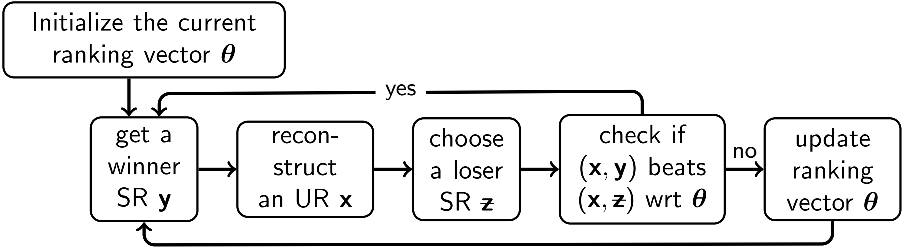
target adult phonotactics

stream of input data

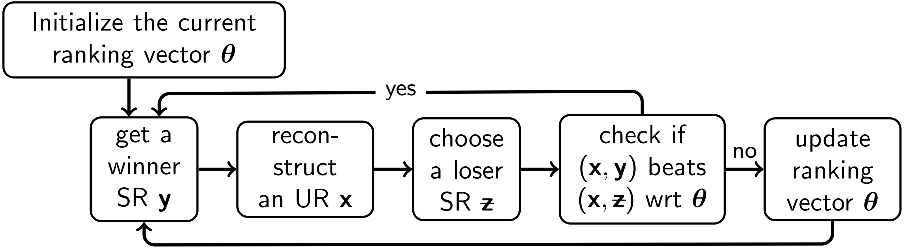
space of all possible grammars

currently predicted phonotactics

EDRA model

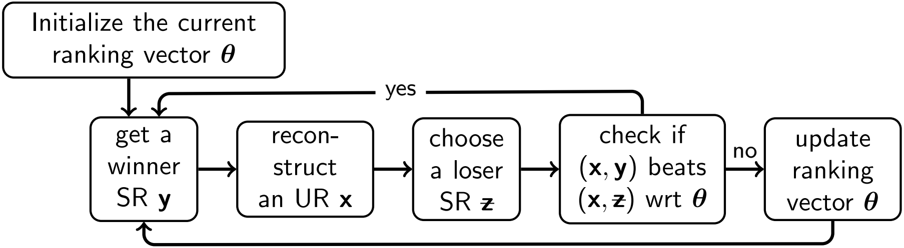

Initialize the current ranking vector $\boldsymbol{\theta}$
yes
get a winner SR $\mathbf{y}$
recon- struct an UR $\mathbf{x}$
choose a loser SR $\mathbf{z}$
check if $(\mathbf{x}, \mathbf{y})$ beats $(\mathbf{x}, \mathbf{z})$ wrt $\boldsymbol{\theta}$
no
update ranking vector $\boldsymbol{\theta}$

EDRA model: representation of the current grammar



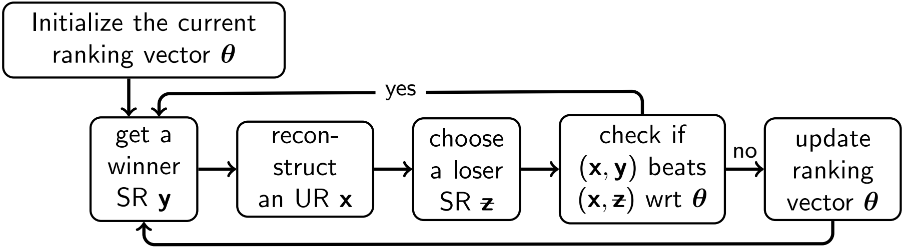
- The algorithm maintains a current hypothesis of the target OT grammar, namely a current constraint ranking
- This current ranking is represented numerically: [Boersma 1998]
  - each constraint is assigned a ranking value
  - big ranking value $\leftrightarrow$ high ranked constraint
  - these ranking values are collected in a ranking vector (RV) θ
- Current RV is initialized and updated through a five step loop

EDRA model: representation of the current grammar



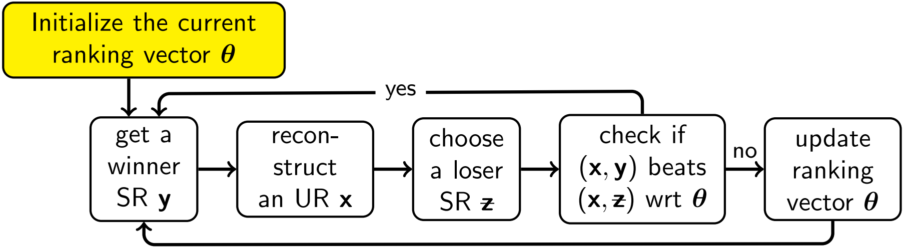
- The algorithm maintains a current hypothesis of the target OT grammar, namely a current constraint ranking
- This current ranking is represented numerically: [Boersma 1998]
  - each constraint is assigned a ranking value
  - big ranking value $\leftrightarrow$ high ranked constraint
  - these ranking values are collected in a ranking vector (RV) $\boldsymbol{\theta}$
- Current RV is initialized and updated through a five step loop

EDRA model: representation of the current grammar

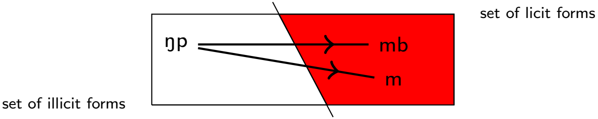


- The algorithm maintains a current hypothesis of the target OT grammar, namely a current constraint ranking
- This current ranking is represented numerically: [Boersma 1998]
  - each constraint is assigned a ranking value
  - big ranking value $\leftrightarrow$ high ranked constraint
  - these ranking values are collected in a ranking vector (RV) $\boldsymbol{\theta}$
- Current RV is initialized and updated through a five step loop

EDRA model: initialization

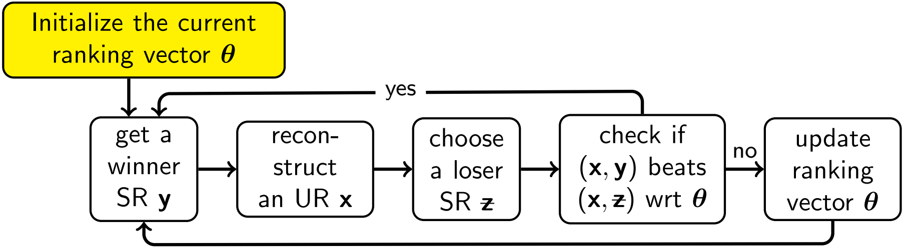


- Optimal repairs are subject to two desiderata:
  - cross the line and always end up with a licit form
  - but do not land too far away from the target

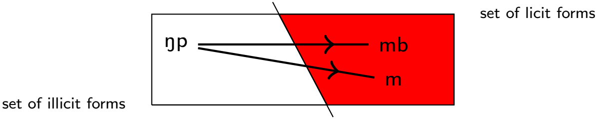


- Two types of OT constraints:
  - markedness: work towards neutralization of contrast (e.g. *NC̥)
  - faithfulness: work towards preservation of contrast (e.g. IDENT$_{\text{voi̥ce}}$)

EDRA model: initialization

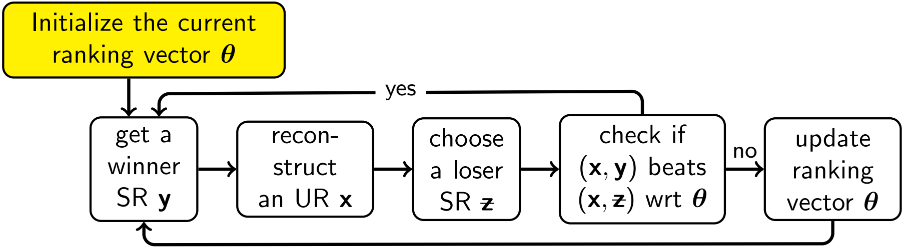


- Optimal repairs are subject to two desiderata:
  - cross the line and always end up with a licit form
  - but do not land too far away from the target



- Two types of OT constraints:
  - markedness: work towards neutralization of contrast (e.g. *NC̥)
  - faithfulness: work towards preservation of contrast (e.g. IDENT$_{\text{voice}}$)

EDRA model: initialization

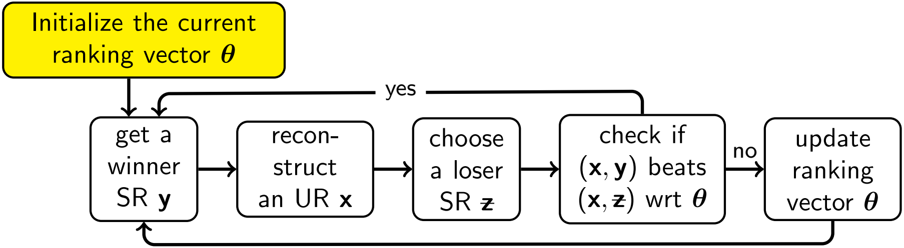


- Initialization of current RV: [Smolensky 1996b,a; Jusczyk, Smolensky, and Allocco 2002]
  - faithfulness constraints start with a small initial ranking value
  - markedness constraints start with a large initial ranking value

$\Rightarrow$ the initial $\mathcal{M} \gg \mathcal{F}$ predicts only unmarked forms to be licit

- The psycholinguistic literature seems to make the opposite assumption! [Davidson, Jusczyk, and Smolensky 2004; Mazuka, Cao, Dupoux, and Christophe 2011]

EDRA model: initialization

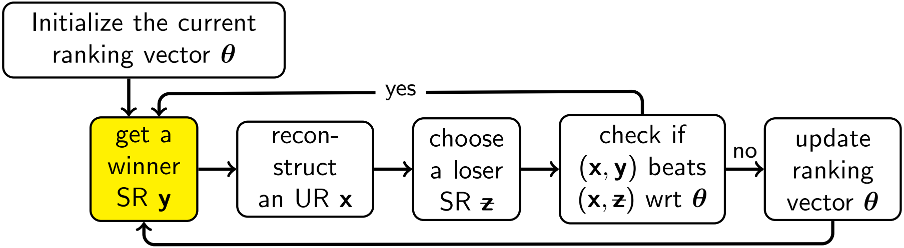


- Initialization of current RV: [Smolensky 1996b,a; Jusczyk, Smolensky, and Allocco 2002]
  - faithfulness constraints start with a small initial ranking value
  - markedness constraints start with a large initial ranking value

$\Rightarrow$ the initial $\mathcal{M} \gg \mathcal{F}$ predicts only unmarked forms to be licit

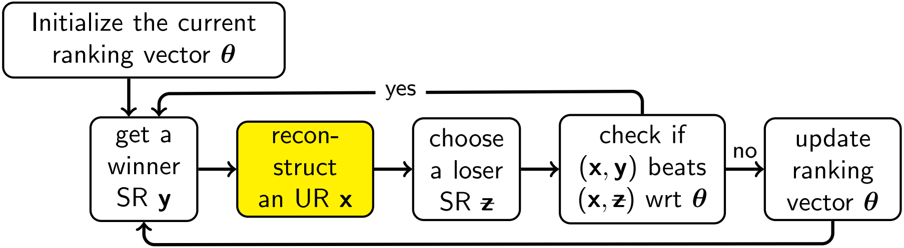
- The psycholinguistic literature seems to make the opposite assumption! [Davidson, Jusczyk, and Smolensky 2004; Mazuka, Cao, Dupoux, and Christophe 2011]

EDRA model: training SR



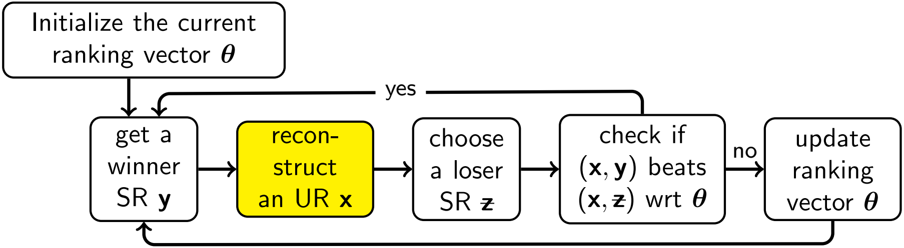
- At each iteration, the model is trained on a SR $\mathbf{y}$ assumed to be licit relative to the target phonotactics
- No assumptions whatsoever are made on this infinite sequence of training data [Cesa-Bianchi and Lugosi 2006]

EDRA model: reconstruction of the UR



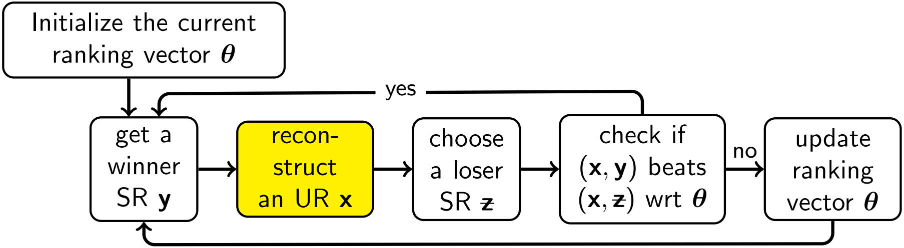
- At each iteration, the model assumes a fully faithful UR $\mathbf{x}$ for the current training SR $\mathbf{y}$: $\boxed{\mathbf{x} = \mathbf{y}}$ [Prince and Tesar 2004]
- This assumption only makes sense if there are no representational differences between SRs and URs [Moreton 2004]
- This assumption is sound if the target grammar is idempotent: does not repair phonotactically licit forms [Magri 2016]
- Idempotency holds "by and large" but not always: chain shifts (a→e→i)

EDRA model: reconstruction of the UR



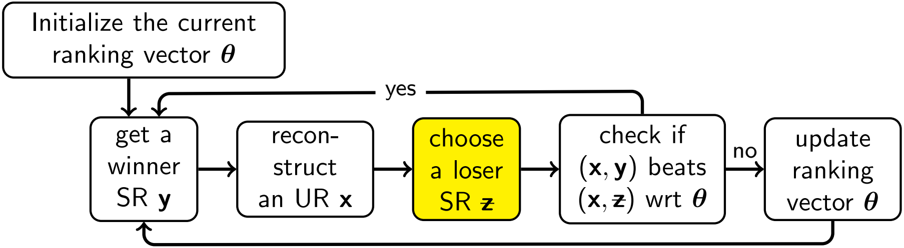
- At each iteration, the model assumes a fully faithful UR **x** for the current training SR **y**: $\boxed{\mathbf{x} = \mathbf{y}}$ [Prince and Tesar 2004]
- This assumption only makes sense if there are no representational differences between SRs and URs [Moreton 2004]
- This assumption is sound if the target grammar is idempotent: does not repair phonotactically licit forms [Magri 2016]
- Idempotency holds "by and large" but not always: chain shifts (a→e→i)

EDRA model: reconstruction of the UR



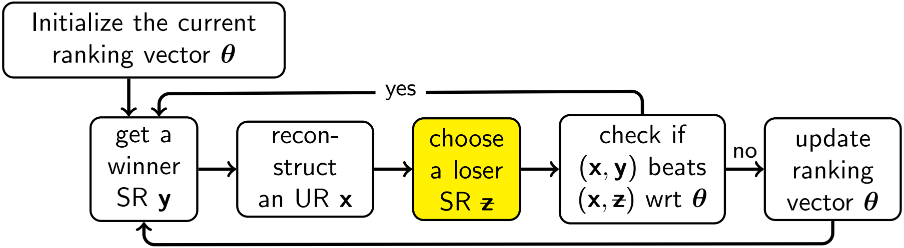
- At each iteration, the model assumes a fully faithful UR $\mathbf{x}$ for the current training SR $\mathbf{y}$: $\boxed{\mathbf{x} = \mathbf{y}}$ [Prince and Tesar 2004]
- This assumption only makes sense if there are no representational differences between SRs and URs [Moreton 2004]
- This assumption is sound if the target grammar is idempotent: does not repair phonotactically licit forms [Magri 2016]
- Idempotency holds "by and large" but not always: chain shifts (a→e→i)

EDRA model: choice of the loser SR



- At each iteration, the model compares the intended winner **y** with some properly chosen loser **z**
- **Rule I:** if there exists a loser SR **z** able to trigger an update, choose one such loser, so as not to waste data [Tesar and Smolensky 1998]
- **Rule II:** If there are multiple such losers, choose a loser SR **z** which is "as close as possible" to the intended winner SR **y** [Magri and Kager 2015]

EDRA model: choice of the loser SR

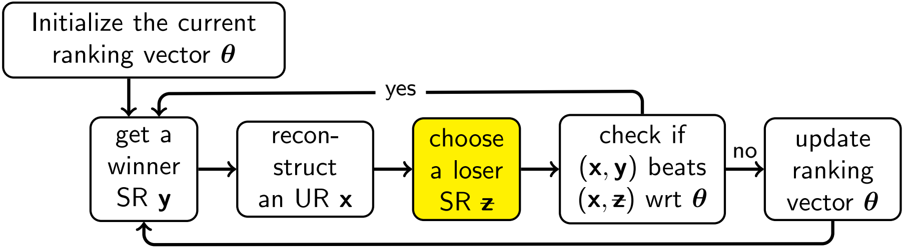


- At each iteration, the model compares the intended winner $\mathbf{y}$ with some properly chosen loser $\mathbf{z}$

- **Rule I:** if there exists a loser SR $\mathbf{z}$ able to trigger an update, choose one such loser, so as not to waste data [Tesar and Smolensky 1998]

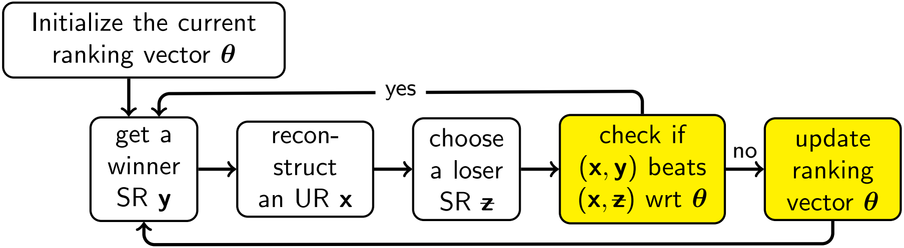
- **Rule II:** If there are multiple such losers, choose a loser SR $\mathbf{z}$ which is "as close as possible" to the intended winner SR $\mathbf{y}$ [Magri and Kager 2015]

EDRA model: choice of the loser SR



- At each iteration, the model compares the intended winner $\mathbf{y}$ with some properly chosen loser $\mathbf{z}$
- **Rule I:** if there exists a loser SR $\mathbf{z}$ able to trigger an update, choose one such loser, so as not to waste data [Tesar and Smolensky 1998]
- **Rule II:** If there are multiple such losers, choose a loser SR $\mathbf{z}$ which is "as close as possible" to the intended winner SR $\mathbf{y}$ [Magri and Kager 2015]

EDRA model: check and update

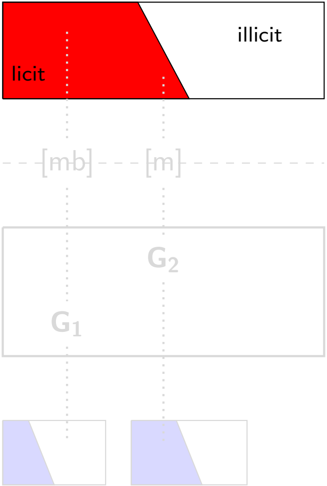


- At each iteration, model checks if the winner mapping $(\mathbf{x}, \mathbf{y})$ beats the loser mapping ($\mathbf{x}$, **ƶ**) according to the current RV
- If that is the case, the learner has nothing to learn from the current piece of data, loops back, and waits for more data
- Otherwise, the learner properly updates the current RV in response to its failure on the current comparison $(\mathbf{x}, \mathbf{y})$ versus ($\mathbf{x}$, **ƶ**)

3.
Convergence, efficiency, consistency

Consistency, convergence, efficiency

- **Learning scenario:**

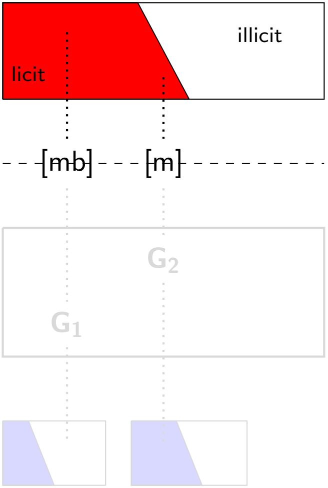


- **Consistency:** Can we guarantee that the set of forms predicted licit is at least as large as the set of target licit forms?

- **Convergence:** equivalently, can we guarantee the EDRA only makes a finite number of errors?
- **Efficiency:** furthermore, can we guarantee the number of errors grows slowly with the number of constraints?

Consistency, convergence, efficiency

- **Learning scenario:**

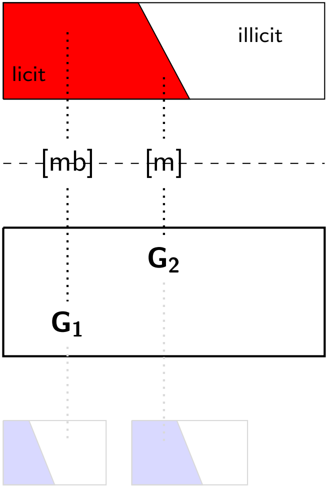


- **Consistency:** Can we guarantee that the set of forms predicted licit is at least as large as the set of target licit forms?

- **Convergence:** equivalently, can we guarantee the EDRA only makes a finite number of errors?
- **Efficiency:** furthermore, can we guarantee the number of errors grows slowly with the number of constraints?

Consistency, convergence, efficiency

- **Learning scenario:**

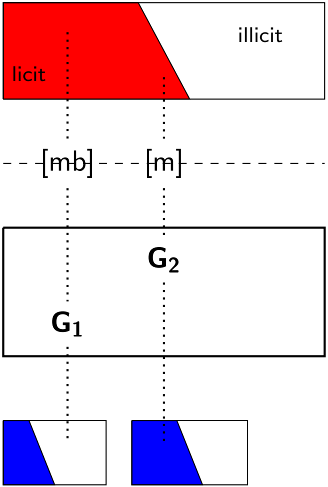


- **Consistency:** Can we guarantee that the set of forms predicted licit is at least as large as the set of target licit forms?

- **Convergence:** equivalently, can we guarantee the EDRA only makes a finite number of errors?
- **Efficiency:** furthermore, can we guarantee the number of errors grows slowly with the number of constraints?

Consistency, convergence, efficiency

- **Learning scenario:**

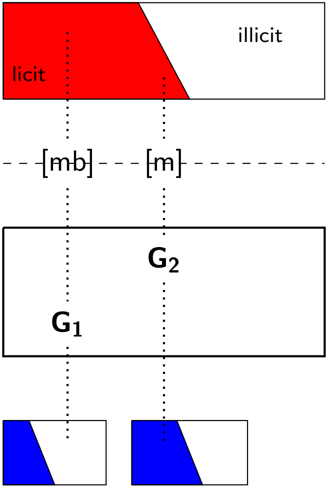


- **Consistency:** Can we guarantee that the set of forms predicted licit is at least as large as the set of target licit forms?

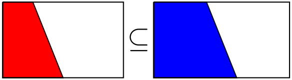
- **Convergence:** equivalently, can we guarantee the EDRA only makes a finite number of errors?
- **Efficiency:** furthermore, can we guarantee the number of errors grows slowly with the number of constraints?

Consistency, convergence, efficiency

- **Learning scenario:**



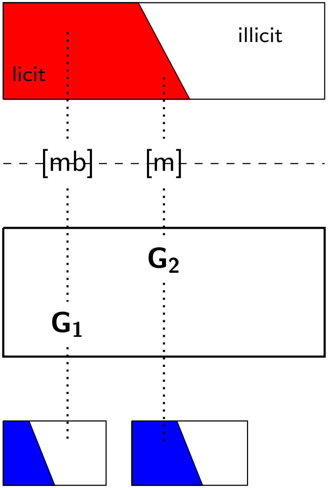
- **Consistency:** Can we guarantee that the set of forms predicted licit is at least as large as the set of target licit forms?



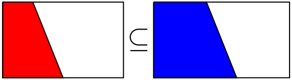
- **Convergence:** equivalently, can we guarantee the EDRA only makes a finite number of errors?
- **Efficiency:** furthermore, can we guarantee the number of errors grows slowly with the number of constraints?

Consistency, convergence, efficiency

- **Learning scenario:**



- **Consistency:** Can we guarantee that the set of forms predicted licit is at least as large as the set of target licit forms?

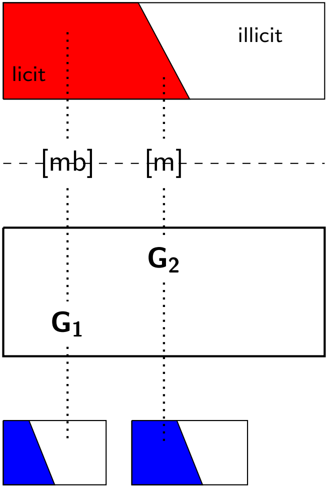


- **Convergence:** equivalently, can we guarantee the EDRA only makes a finite number of errors?

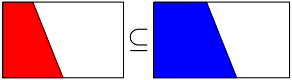
- **Efficiency:** furthermore, can we guarantee the number of errors grows slowly with the number of constraints?

Consistency, convergence, efficiency

- **Learning scenario:**

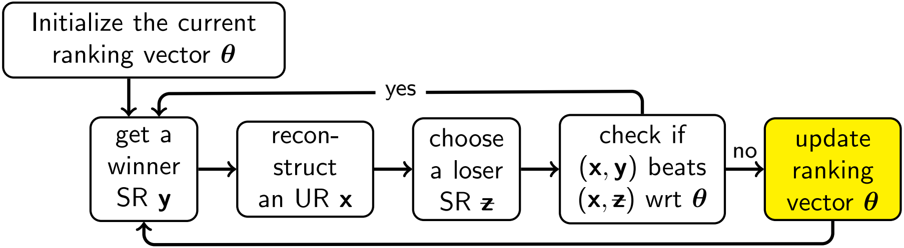


- **Consistency:** Can we guarantee that the set of forms predicted licit is at least as large as the set of target licit forms?



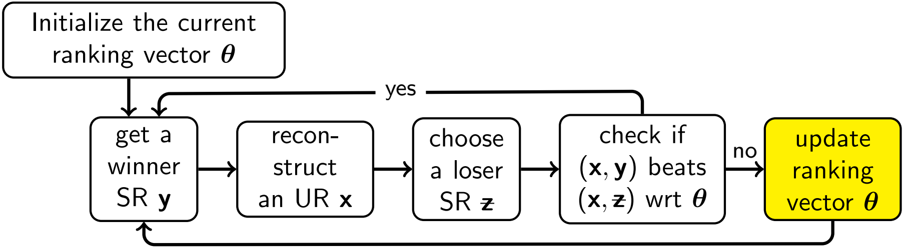
- **Convergence:** equivalently, can we guarantee the EDRA only makes a finite number of errors?
- **Efficiency:** furthermore, can we guarantee the number of errors grows slowly with the number of constraints?

How to tune the promotion amount to get convergence



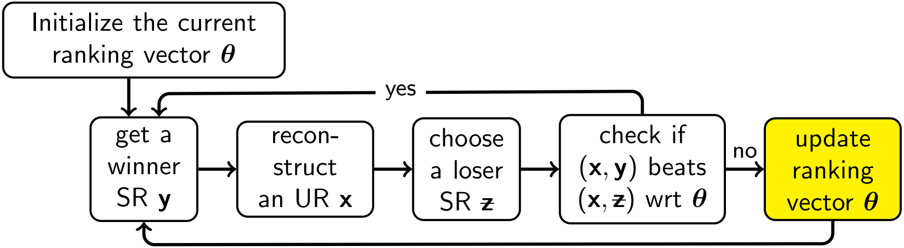
- **Demotion component:**
  - decrease ranking value of (undominated) loser-preferring constraints
  - by a certain amount, say 1 for concreteness
- **Promotion component:**
  - increase ranking value of winner-preferring constraints
  - by a certain promotion amount, call it $p \geq 0$
- **Examples:**
  - $p = 0$: (gradual) EDCD [Tesar and Smolensky 1998]
  - $p = 1$: (non-stochatic) GLA [Boersma 1997, 1998]
- **Question:** how to choose p so that we get efficient convergence?

How to tune the promotion amount to get convergence



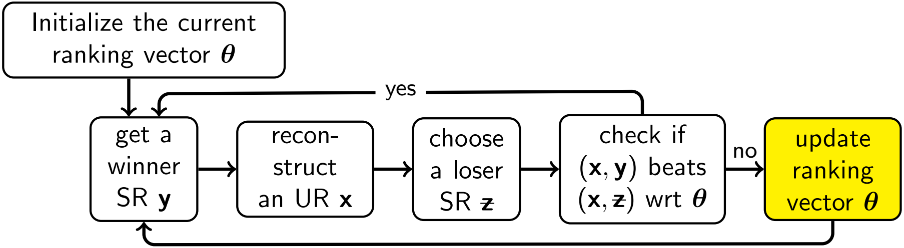
- **Demotion component:**
  - decrease ranking value of (undominated) loser-preferring constraints
  - by a certain amount, say 1 for concreteness
- **Promotion component:**
  - increase ranking value of winner-preferring constraints
  - by a certain promotion amount, call it $p \geq 0$
- **Examples:**
  - $p = 0$: (gradual) EDCD [Tesar and Smolensky 1998]
  - $p = 1$: (non-stochatic) GLA [Boersma 1997, 1998]
- **Question:** how to choose p so that we get efficient convergence?

How to tune the promotion amount to get convergence



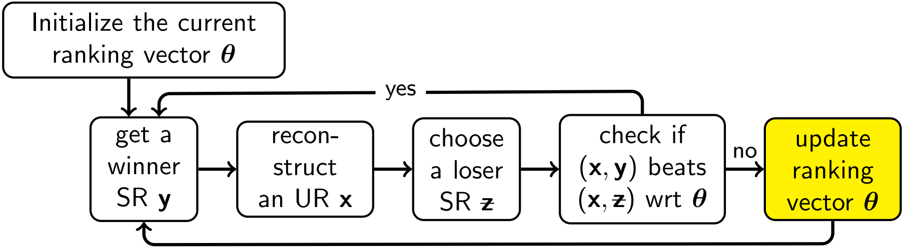
- **Demotion component:**
  - decrease ranking value of (undominated) loser-preferring constraints
  - by a certain amount, say 1 for concreteness
- **Promotion component:**
  - increase ranking value of winner-preferring constraints
  - by a certain promotion amount, call it $p \geq 0$
- **Examples:**
  - $p = 0$: (gradual) EDCD [Tesar and Smolensky 1998]
  - $p = 1$: (non-stochatic) GLA [Boersma 1997, 1998]
- **Question:** how to choose p so that we get efficient convergence?

How to tune the promotion amount to get convergence



- **Demotion component:**
  - decrease ranking value of (undominated) loser-preferring constraints
  - by a certain amount, say 1 for concreteness
- **Promotion component:**
  - increase ranking value of winner-preferring constraints
  - by a certain promotion amount, call it $p \geq 0$
- **Examples:**
  - $p = 0$: (gradual) EDCD [Tesar and Smolensky 1998]
  - $p = 1$: (non-stochatic) GLA [Boersma 1997, 1998]
- **Question:** how to choose p so that we get efficient convergence?

How to tune the promotion amount to get convergence

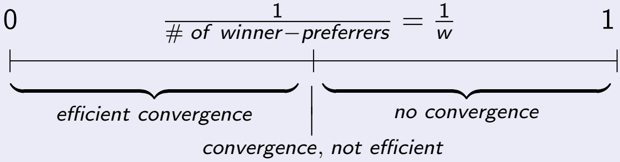


- **Demotion component:**
  - decrease ranking value of (undominated) loser-preferring constraints
  - by a certain amount, say 1 for concreteness
- **Promotion component:**
  - increase ranking value of winner-preferring constraints
  - by a certain promotion amount, call it $p \geq 0$
- **Examples:**
  - $p = 0$: (gradual) EDCD [Tesar and Smolensky 1998]
  - $p = 1$: (non-stochatic) GLA [Boersma 1997, 1998]
- **Question:** how to choose p so that we get efficient convergence?

A complete theory of convergence

First result of this talk

Efficient convergence holds iff the promotion amount p is calibrated namely smaller than the inverse of the number of winner-preferrers:



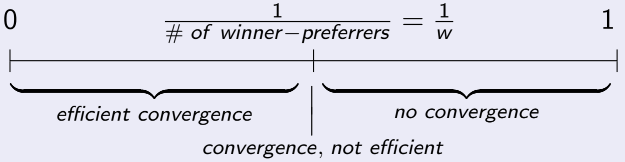
[Tesar and Smolensky 1998; Pater 2008; Magri 2012]

- Some boasting:
  - complete theory: necessary and sufficient
  - no assumptions on initialization, URs, losers
  - extends to noisy setting [Magri to appearb]
  - as well as to the stochastic implementation used for variation [Boersma 1998]

A complete theory of convergence

First result of this talk

Efficient convergence holds iff the promotion amount p is calibrated namely smaller than the inverse of the number of winner-preferrers:



[Tesar and Smolensky 1998; Pater 2008; Magri 2012]

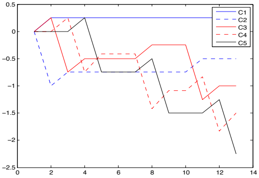
- Some boasting:
  - complete theory: necessary and sufficient
  - no assumptions on initialization, URs, losers
  - extends to noisy setting [Magri to appearb]
  - as well as to the stochastic implementation used for variation [Boersma 1998]

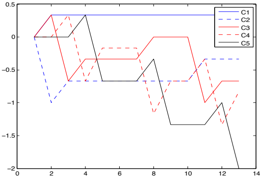
Discussion/I

- Expect the model to behave similarly for similar values of p

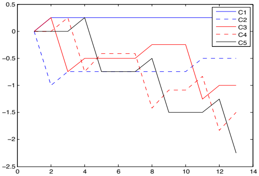
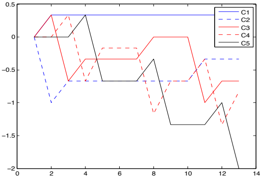
Discussion/I

- Expect the model to behave similarly for similar values of p

- $p = \frac{1}{w+2}$ 

$p = \frac{1}{w+1}$ 

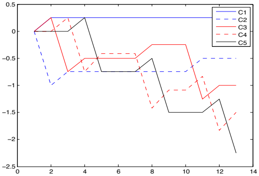
Discussion/I

- Expect the model to behave similarly for similar values of p
- $p = \frac{1}{w+2}$  $p = \frac{1}{w+1}$ 
- $p = \frac{1}{w+0}$

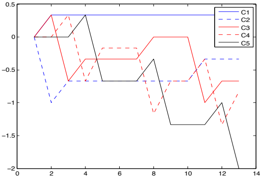
Discussion/I

- Expect the model to behave similarly for similar values of p

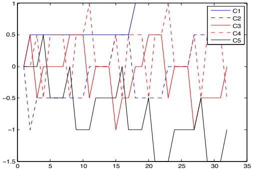
- $p = \frac{1}{w+2}$



$p = \frac{1}{w+1}$

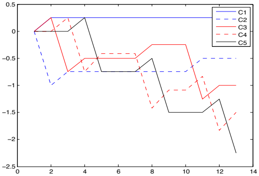


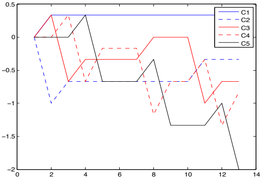
- $p = \frac{1}{w+0}$

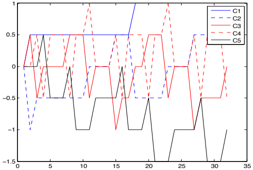


- $p = \frac{1}{w}$ is a breaking point for both convergence and modeling

Discussion/I

- Expect the model to behave similarly for similar values of p
- $p = \frac{1}{w+2}$ 

  $p = \frac{1}{w+1}$ 

- $p = \frac{1}{w+0}$ 

- $p = \frac{1}{w}$ is a breaking point for both convergence and modeling

Discussion/II

- "That strict domination governs grammatical constraint interaction is not currently explained" [Prince and Smolensky 1997]
- "a [. . .] possibility is that demands of learnability provide a pressure for strict domination among constraints, [although] it remains an open problem to formally characterize exactly what is essential about strict domination to guarantee efficient learning" [Legendre et al. 2006]
- Efficient convergence in OT can be guaranteed with no assumptions on the constraints: it follows from OT's "tilted planes"
- And it does not extend beyond it: [Magri to appeara]

Discussion/II

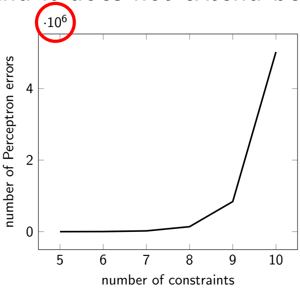
- "That strict domination governs grammatical constraint interaction is not currently explained" [Prince and Smolensky 1997]
- "a [. . .] possibility is that demands of learnability provide a pressure for strict domination among constraints, [although] it remains an open problem to formally characterize exactly what is essential about strict domination to guarantee efficient learning" [Legendre et al. 2006]
- Efficient convergence in OT can be guaranteed with no assumptions on the constraints: it follows from OT's "tilted planes"
- And it does not extend beyond it: [Magri to appeara]

Discussion/II

- "That strict domination governs grammatical constraint interaction is not currently explained" [Prince and Smolensky 1997]
- "a [. . .] possibility is that demands of learnability provide a pressure for strict domination among constraints, [although] it remains an open problem to formally characterize exactly what is essential about strict domination to guarantee efficient learning" [Legendre et al. 2006]
- Efficient convergence in OT can be guaranteed with no assumptions on the constraints: it follows from OT's "tilted planes"
- And it does not extend beyond it: [Magri to appeara]

Discussion/II

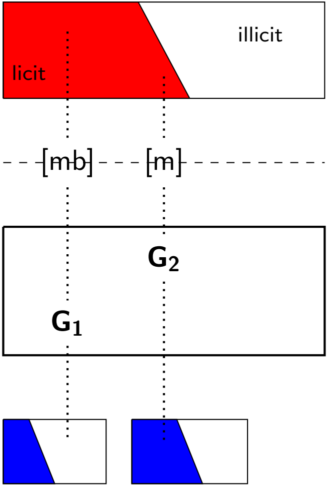
- "That strict domination governs grammatical constraint interaction is not currently explained" [Prince and Smolensky 1997]
- "a [. . .] possibility is that demands of learnability provide a pressure for strict domination among constraints, [although] it remains an open problem to formally characterize exactly what is essential about strict domination to guarantee efficient learning" [Legendre et al. 2006]
- Efficient convergence in OT can be guaranteed with no assumptions on the constraints: it follows from OT's "tilted planes"
- And it does not extend beyond it: [Magri to appeara]



4.
Restrictiveness

Restrictiveness

- **Learning scenario:**



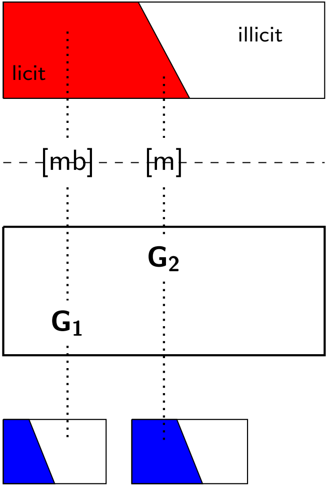
- **Restrictiveness:** Can we guarantee that at every iteration the set of forms predicted licit is not larger than the set of target licit forms?

- **Convergence** (⊆)
  \+ **Restrictiveness** (⊇)
  = **Correctness**

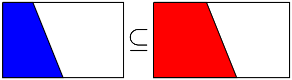
- Unfortunately, restrictiveness is a much trickier business than convergence!

Restrictiveness

- **Learning scenario:**



- **Restrictiveness:** Can we guarantee that at every iteration the set of forms predicted licit is not larger than the set of target licit forms?

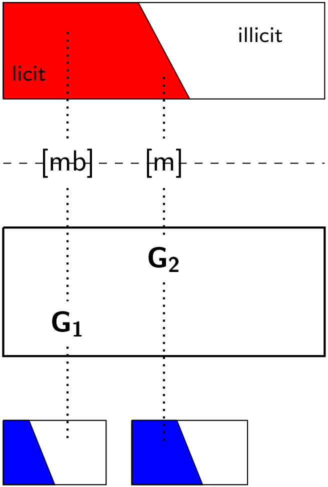


- **Convergence** (⊆)
  \+ **Restrictiveness** (⊇)
  = **Correctness**

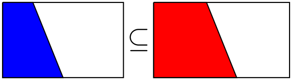
- Unfortunately, restrictiveness is a much trickier business than convergence!

Restrictiveness

- **Learning scenario:**



- **Restrictiveness:** Can we guarantee that at every iteration the set of forms predicted licit is not larger than the set of target licit forms?

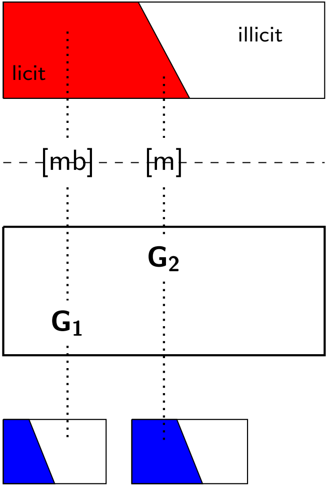


- $$\begin{array}{ll} & \textbf{Convergence } (\subseteq) \\ + & \textbf{Restrictiveness } (\supseteq) \\ \hline = & \textbf{Correctness} \end{array}$$

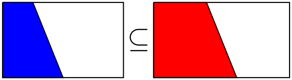
- Unfortunately, restrictiveness is a much trickier business than convergence!

Restrictiveness

- **Learning scenario:**



- **Restrictiveness:** Can we guarantee that at every iteration the set of forms predicted licit is not larger than the set of target licit forms?



- **Convergence** (⊆)
  \+ **Restrictiveness** (⊇)
  = **Correctness**

- Unfortunately, restrictiveness is a much trickier business than convergence!

The modeling problem is intractable

- **Given:** [Fodor and Sakas 2005]
  - a finite OT typology specified through candidates and constraints
  - a finite set of data consisting of *consistent* URs/SRs pairs
- **Find:** a ranking over the constraint set which is
  - consistent: the corresponding grammar enforces the training mappings
  - restrictive: there is no other consistent ranking which yields a more restrictive phonotactics
- **Size:** $\max\{\text{number of constraints}, \text{number of candidates}\}$

Second result of this talk

This formulation of the Subset problem in OT is intractable [Magri 2013]

— despite the algorithm being allowed to list all candidates

— even if the problem is restricted to the simplest disjunctive structure

(Reduction from CYCLICORDERINGPROBLEM). [Galil and Megiddo 1977]

The modeling problem is intractable

- **Given:** [Fodor and Sakas 2005]
  - a finite OT typology specified through candidates and constraints
  - a finite set of data consisting of *consistent* URs/SRs pairs
- **Find:** a ranking over the constraint set which is
  - consistent: the corresponding grammar enforces the training mappings
  - restrictive: there is no other consistent ranking which yields a more restrictive phonotactics
- **Size:** max $\{$number of constraints, number of candidates$\}$

Second result of this talk

This formulation of the Subset problem in OT is intractable [Magri 2013]

— despite the algorithm being allowed to list all candidates

— even if the problem is restricted to the simplest disjunctive structure

(Reduction from CyclicOrderingProblem). [Galil and Megiddo 1977]

The modeling problem is intractable

- **Given:** [Fodor and Sakas 2005]
  - a finite OT typology specified through candidates and constraints
  - a finite set of data consisting of *consistent* URs/SRs pairs
- **Find:** a ranking over the constraint set which is
  - consistent: the corresponding grammar enforces the training mappings
  - restrictive: there is no other consistent ranking which yields a more restrictive phonotactics
- **Size:** $\max\{\text{number of constraints}, \text{number of candidates}\}$

Second result of this talk

This formulation of the Subset problem in OT is intractable [Magri 2013]

— despite the algorithm being allowed to list all candidates

— even if the problem is restricted to the simplest disjunctive structure

(Reduction from CYCLICORDERINGPROBLEM). [Galil and Megiddo 1977]

The modeling problem is intractable

- **Given:** [Fodor and Sakas 2005]
  - a finite OT typology specified through candidates and constraints
  - a finite set of data consisting of *consistent* URs/SRs pairs
- **Find:** a ranking over the constraint set which is
  - consistent: the corresponding grammar enforces the training mappings
  - restrictive: there is no other consistent ranking which yields a more restrictive phonotactics
- **Size:** $\max\{\text{number of constraints}, \text{number of candidates}\}$

Second result of this talk

This formulation of the Subset problem in OT is intractable [Magri 2013]

— despite the algorithm being allowed to list all candidates

— even if the problem is restricted to the simplest disjunctive structure

(Reduction from CyclicOrderingProblem). [Galil and Megiddo 1977]

Discussion

- **Not specific to OT:**
  - "complexity problems are largely independent of the learning paradigm, as all frameworks encounter them" [Clark and Lappin 2011]
  - intractability result different from previous OT-specific ones [Idsardi 2006]
- **Restrictiveness requires assumptions:** [Barton, Berwick, and Ristad 1987]
  - "[intractable] pbms don't have any special structure that would support an efficient solution algorithm, so there's little choice but brute force"
  - they "might well be characterized as unnatural": "there is every reason to believe that natural language has an intricate computational structure that is not reflected in combinatorial search methods"
  - they "leave unmentioned some constraints of the natural problem"

  If the child solves it, the problem has got to be solvable!
- **Two possible types of assumptions:**
  - restrictiveness through assumptions on the target pattern, for any underlying constraint set
  - restrictiveness through assumptions on the constraint set, for any target pattern in the corresponding typology

Discussion

- **Not specific to OT:**
  - "complexity problems are largely independent of the learning paradigm, as all frameworks encounter them" [Clark and Lappin 2011]
  - intractability result different from previous OT-specific ones [Idsardi 2006]
- **Restrictiveness requires assumptions:** [Barton, Berwick, and Ristad 1987]
  - "[intractable] pbms don't have any special structure that would support an efficient solution algorithm, so there's little choice but brute force"
  - they "might well be characterized as unnatural": "there is every reason to believe that natural language has an intricate computational structure that is not reflected in combinatorial search methods"
  - they "leave unmentioned some constraints of the natural problem"

If the child solves it, the problem has got to be solvable!

- **Two possible types of assumptions:**
  - restrictiveness through assumptions on the target pattern, for any underlying constraint set
  - restrictiveness through assumptions on the constraint set, for any target pattern in the corresponding typology

Discussion

- **Not specific to OT:**
  - "complexity problems are largely independent of the learning paradigm, as all frameworks encounter them" [Clark and Lappin 2011]
  - intractability result different from previous OT-specific ones [Idsardi 2006]
- **Restrictiveness requires assumptions:** [Barton, Berwick, and Ristad 1987]
  - "[intractable] pbms don't have any special structure that would support an efficient solution algorithm, so there's little choice but brute force"
  - they "might well be characterized as unnatural": "there is every reason to believe that natural language has an intricate computational structure that is not reflected in combinatorial search methods"
  - they "leave unmentioned some constraints of the natural problem"

If the child solves it, the problem has got to be solvable!

- **Two possible types of assumptions:**
  - restrictiveness through assumptions on the target pattern, for any underlying constraint set
  - restrictiveness through assumptions on the constraint set, for any target pattern in the corresponding typology

Discussion

- **Not specific to OT:**
  - "complexity problems are largely independent of the learning paradigm, as all frameworks encounter them" [Clark and Lappin 2011]
  - intractability result different from previous OT-specific ones [Idsardi 2006]
- **Restrictiveness requires assumptions:** [Barton, Berwick, and Ristad 1987]
  - "[intractable] pbms don't have any special structure that would support an efficient solution algorithm, so there's little choice but brute force"
  - they "might well be characterized as unnatural": "there is every reason to believe that natural language has an intricate computational structure that is not reflected in combinatorial search methods"
  - they "leave unmentioned some constraints of the natural problem"

  If the child solves it, the problem has got to be solvable!
- **Two possible types of assumptions:**
  - restrictiveness through assumptions on the target pattern, for any underlying constraint set
  - restrictiveness through assumptions on the constraint set, for any target pattern in the corresponding typology

Discussion

- **Not specific to OT:**
  - "complexity problems are largely independent of the learning paradigm, as all frameworks encounter them" [Clark and Lappin 2011]
  - intractability result different from previous OT-specific ones [Idsardi 2006]
- **Restrictiveness requires assumptions:** [Barton, Berwick, and Ristad 1987]
  - "[intractable] pbms don't have any special structure that would support an efficient solution algorithm, so there's little choice but brute force"
  - they "might well be characterized as unnatural": "there is every reason to believe that natural language has an intricate computational structure that is not reflected in combinatorial search methods"
  - they "leave unmentioned some constraints of the natural problem"

If the child solves it, the problem has got to be solvable!

- **Two possible types of assumptions:**
  - restrictiveness through assumptions on the target pattern, for any underlying constraint set
  - restrictiveness through assumptions on the constraint set, for any target pattern in the corresponding typology

Faithfulness constraints and phonotactics

- **Velar inventories:**
  - let's focus of segment inventories
  - out of the four velar obstruents: [g], [k], [ɣ], [x]
  - here are some representative examples: [g k ~~ɣ~~ ~~x~~] [~~g~~ k ~~ɣ~~ ~~x~~]

- **Phonology:** relative ranking of $\mathcal{F}$-constraints does matter

IDENT$_{[place]}$
|
IDENT$_{[cont]}$
⇒ /ɣ/ → [z]

IDENT$_{[cont]}$
|
IDENT$_{[place]}$
⇒ /ɣ/ → [g]

- **Phonotactics:** relative ranking of $\mathcal{F}$-constraints does not matter

NoDorFric
|
all $\mathcal{F}$ constraints
|
remaining $\mathcal{M}$ constraints

NoDorFric, **NoVoiStop**
|
all $\mathcal{F}$ constraints
|
remaining $\mathcal{M}$ constraints

Faithfulness constraints and phonotactics

- **Velar inventories:**
  - let's focus of segment inventories
  - out of the four velar obstruents: [g], [k], [ɣ], [x]
  - here are some representative examples: [g k ~~ɣ~~ ~~x~~] [~~g~~ k ~~ɣ~~ ~~x~~]

- **Phonology:** relative ranking of $\mathcal{F}$-constraints does matter

$$\begin{array}{c} \text{IDENT}_{[place]} \\ | \\ \text{IDENT}_{[cont]} \end{array} \Rightarrow /\text{ɣ}/ \rightarrow [\text{z}] \qquad \begin{array}{c} \text{IDENT}_{[cont]} \\ | \\ \text{IDENT}_{[place]} \end{array} \Rightarrow /\text{ɣ}/ \rightarrow [\text{g}]$$

- **Phonotactics:** relative ranking of $\mathcal{F}$-constraints does not matter

$$\begin{array}{c} \textbf{NoDorFric} \\ | \\ \text{all } \mathcal{F} \text{ constraints} \\ | \\ \text{remaining } \mathcal{M} \text{ constraints} \end{array} \qquad \begin{array}{c} \textbf{NoDorFric}, \textbf{NoVoiStop} \\ | \\ \text{all } \mathcal{F} \text{ constraints} \\ | \\ \text{remaining } \mathcal{M} \text{ constraints} \end{array}$$

Faithfulness constraints and phonotactics

- **Velar inventories:**
  - let's focus of segment inventories
  - out of the four velar obstruents: [g], [k], [ɣ], [x]
  - here are some representative examples: [g k ~~ɣ~~ ~~x~~] [~~g~~ k ~~ɣ~~ ~~x~~]

- **Phonology:** relative ranking of $\mathcal{F}$-constraints does matter

$\textsc{Ident}_{[place]}$ | $\textsc{Ident}_{[cont]}$ ⇒ /ɣ/ → [z]

$\textsc{Ident}_{[cont]}$ | $\textsc{Ident}_{[place]}$ ⇒ /ɣ/ → [g]

- **Phonotactics:** relative ranking of $\mathcal{F}$-constraints does not matter

NoDorFric
|
all $\mathcal{F}$ constraints
|
remaining $\mathcal{M}$ constraints

NoDorFric, **NoVoiStop**
|
all $\mathcal{F}$ constraints
|
remaining $\mathcal{M}$ constraints

An assumption on the target phonotactic pattern

- **Intuition:**
  - OT constraints come in two varieties: $\mathcal{F}$ and $\mathcal{M}$ constraints
  - relative ranking of $\mathcal{F}$ constraints determines how illicit forms are repaired
  - but it contributes little to the distinction between licit/illicit forms
  - namely, it is irrelevant for phonotactics

- **Formalization:**

A phonotactic pattern is $\mathcal{F}$-irrelevant (relative to a certain constraint set) provided it can be generated by

$$\mathcal{M}_{top} \gg \mathcal{F} \gg \mathcal{M}_{bottom}$$

for some partition $\mathcal{M}_{top}, \mathcal{M}_{bottom}$ of the markedness constraints

- **How general is this assumption?**
  - some authors have (implicitly) assumed it is universal [Hayes and Wilson 2008]
  - it really depends on the theory of markedness:
    - — phonetic grounding [Hayes and Steriade 2004]
    - — positional faitthfulness [Beckman 1999]

An assumption on the target phonotactic pattern

- **Intuition:**
  - OT constraints come in two varieties: $\mathcal{F}$ and $\mathcal{M}$ constraints
  - relative ranking of $\mathcal{F}$ constraints determines how illicit forms are repaired
  - but it contributes little to the distinction between licit/illicit forms
  - namely, it is irrelevant for phonotactics

- **Formalization:**

A phonotactic pattern is $\mathcal{F}$-irrelevant (relative to a certain constraint set) provided it can be generated by

$$\mathcal{M}_{top} \gg \mathcal{F} \gg \mathcal{M}_{bottom}$$

for some partition $\mathcal{M}_{top}, \mathcal{M}_{bottom}$ of the markedness constraints

- **How general is this assumption?**
  - some authors have (implicitly) assumed it is universal [Hayes and Wilson 2008]
  - it really depends on the theory of markedness:
    - — phonetic grounding [Hayes and Steriade 2004]
    - — positional faitthfulness [Beckman 1999]

An assumption on the target phonotactic pattern

Intuition:

- OT constraints come in two varieties: $\mathcal{F}$ and $\mathcal{M}$ constraints
- relative ranking of $\mathcal{F}$ constraints determines how illicit forms are repaired
- but it contributes little to the distinction between licit/illicit forms
- namely, it is irrelevant for phonotactics

Formalization:

A phonotactic pattern is $\mathcal{F}$-irrelevant (relative to a certain constraint set) provided it can be generated by

$$\mathcal{M}_{top} \gg \mathcal{F} \gg \mathcal{M}_{bottom}$$

for some partition $\mathcal{M}_{top}, \mathcal{M}_{bottom}$ of the markedness constraints

How general is this assumption?

- some authors have (implicitly) assumed it is universal [Hayes and Wilson 2008]
- it really depends on the theory of markedness:
  - — phonetic grounding [Hayes and Steriade 2004]
  - — positional faitthfulness [Beckman 1999]

EDRAs' restrictiveness on $\mathcal{F}$-irrelevant languages

Third result of this talk

The EDRA model is restrictive provided: [Magri in preparation]

Hp1 the target/training phonotactic pattern is $\mathcal{F}$-irrelevant

EDRAs' restrictiveness on $\mathcal{F}$-irrelevant languages

Third result of this talk

The EDRA model is restrictive provided: [Magri in preparation]

Hp1 the target/training phonotactic pattern is $\mathcal{F}$-irrelevant

Hp2 $p \leq \frac{1}{w + \Delta}$ with $\Delta(m) \simeq \frac{m}{k \log m}$

Hp3 the candidacy relation is symmetric

Hp4 faithfulness constraints are distinctive

- Mild boasting:
  - a restrictiveness result for at least a large majority of patterns
  - virtually no assumptions on the constraints (Hp4 very mild)
  - no assumptions on the subroutine for the choice of the loser

EDRAs' restrictiveness on $\mathcal{F}$-irrelevant languages

Third result of this talk

The EDRA model is restrictive provided: [Magri in preparation]

- Hp1 the target/training phonotactic pattern is $\mathcal{F}$-irrelevant
- Hp2 $p \leq \frac{1}{w + \Delta}$ with $\Delta(m) \simeq \frac{m}{k \log m}$
- Hp3 the candidacy relation is symmetric
- Hp4 faithfulness constraints are distinctive

□ Mild boasting:

- a restrictiveness result for at least a large majority of patterns
- virtually no assumptions on the constraints (Hp4 very mild)
- no assumptions on the subroutine for the choice of the loser

EDRAs' restrictiveness on $\mathcal{F}$-irrelevant languages

Third result of this talk

The EDRA model is restrictive provided: [Magri in preparation]

Hp1 the target/training phonotactic pattern is $\mathcal{F}$-irrelevant

Hp2 $p \leq \dfrac{1}{w + \Delta}$ with $\Delta(m) \simeq \dfrac{m}{k \log m}$

Hp3 the candidacy relation is symmetric

Hp4 faithfulness constraints are distinctive

- Faithfulness constraints militate in favor of contrast preservation
- restrictiveness thus requires them to be ranked as low as possible
- i.e., to be promoted little from their initial low ranking position
- Hp2 indeed requires the promotion amount to be small

EDRAs' restrictiveness on $\mathcal{F}$-irrelevant languages

Third result of this talk

The EDRA model is restrictive provided: [Magri in preparation]

- **Hp1** the target/training phonotactic pattern is $\mathcal{F}$-irrelevant
- **Hp2** $p \leq \frac{1}{w + \Delta}$ with $\Delta(m) \simeq \frac{m}{k \log m}$
- **Hp3** the candidacy relation is symmetric
- **Hp4** faithfulness constraints are distinctive

- Faithfulness constraints militate in favor of contrast preservation
- restrictiveness thus requires them to be ranked as low as possible
- i.e., to be promoted little from their initial low ranking position
- Hp2 indeed requires the promotion amount to be small

EDRAs' restrictiveness on $\mathcal{F}$-irrelevant languages

Third result of this talk

The EDRA model is restrictive provided: [Magri in preparation]

- **Hp1** the target/training phonotactic pattern is $\mathcal{F}$-irrelevant
- **Hp2** $p \leq \frac{1}{w+\Delta}$ with $\Delta(m) \simeq \frac{m}{k \log m}$
- **Hp3** the candidacy relation is symmetric
- **Hp4** faithfulness constraints are distinctive

- Suppose **x** is a target illicit phonological form
- **x** is neutralized to some licit candidate form **y** by target grammar

EDRAs' restrictiveness on $\mathcal{F}$-irrelevant languages

Third result of this talk

The EDRA model is restrictive provided: [Magri in preparation]

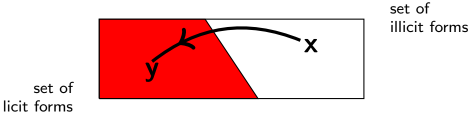
Hp1 the target/training phonotactic pattern is $\mathcal{F}$-irrelevant

Hp2 $p \leq \frac{1}{w + \Delta}$ with $\Delta(m) \simeq \frac{m}{k \log m}$

Hp3 the candidacy relation is symmetric

Hp4 faithfulness constraints are distinctive

- Suppose **x** is a target illicit phonological form
- **x** is neutralized to some licit candidate form **y** by target grammar



EDRAs' restrictiveness on $\mathcal{F}$-irrelevant languages

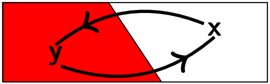
Third result of this talk

The EDRA model is restrictive provided: [Magri in preparation]

- **Hp1** the target/training phonotactic pattern is $\mathcal{F}$-irrelevant
- **Hp2** $p \leq \frac{1}{w + \Delta}$ with $\Delta(m) \simeq \frac{m}{k \log m}$
- **Hp3** the candidacy relation is symmetric
- **Hp4** faithfulness constraints are distinctive

- ☐ Suppose **x** is a target illicit phonological form
- ☐ **x** is neutralized to some licit candidate form **y** by target grammar

set of illicit forms



set of licit forms

- ☐ symmetry requires that vice versa **x** be a candidate of **y**
- ☐ Plausible if candidacy defined in terms of phonological operations

EDRAs' restrictiveness on $\mathcal{F}$-irrelevant languages

Third result of this talk

The EDRA model is restrictive provided: [Magri in preparation]

- **Hp1** the target/training phonotactic pattern is $\mathcal{F}$-irrelevant
- **Hp2** $p \leq \frac{1}{w + \Delta}$ with $\Delta(m) \simeq \frac{m}{k \log m}$
- **Hp3** the candidacy relation is symmetric
- **Hp4** faithfulness constraints are distinctive

- In order for the $\mathcal{F}$-irrelevance assumption to bite, there have got to be enough faithfulness constraints
- faithfulness distinciveness ensures precisely that
- as it requires that any two candidate SRs $\mathbf{y}, \mathbf{z}$ for the same UR $\mathbf{x}$ be distinguished by at least a faithfulness constraint: $F(\mathbf{x}, \mathbf{y}) \neq F(\mathbf{x}, \mathbf{z})$
- Plausible assumption to connect candidates and constraints

EDRAs' restrictiveness on $\mathcal{F}$-irrelevant languages

Third result of this talk

The EDRA model is restrictive provided: [Magri in preparation]

- **Hp1** the target/training phonotactic pattern is $\mathcal{F}$-irrelevant
- **Hp2** $p \leq \frac{1}{w + \Delta}$ with $\Delta(m) \simeq \frac{m}{k \log m}$
- **Hp3** the candidacy relation is symmetric
- **Hp4** faithfulness constraints are distinctive

- In order for the $\mathcal{F}$-irrelevance assumption to bite, there have got to be enough faithfulness constraints
- faithfulness distinciveness ensures precisely that
- as it requires that any two candidate SRs $\mathbf{y}, \mathbf{z}$ for the same UR $\mathbf{x}$ be distinguished by at least a faithfulness constraint: $\boxed{F(\mathbf{x}, \mathbf{y}) \neq F(\mathbf{x}, \mathbf{z})}$
- Plausible assumption to connect candidates and constraints

EDRAs' restrictiveness on $\mathcal{F}$-irrelevant languages

Third result of this talk

The EDRA model is restrictive provided: [Magri in preparation]

- Hp1 the target/training phonotactic pattern is $\mathcal{F}$-irrelevant
- Hp2 $p \leq \frac{1}{w + \Delta}$ with $\Delta(m) \simeq \frac{m}{k \log m}$
- Hp3 the candidacy relation is symmetric
- Hp4 faithfulness constraints are distinctive

- Idea of the proof:
  - By Hp1, the target ranking is $\mathcal{M}_{top} \gg \mathcal{F} \gg \mathcal{M}_{bottom}$
  - By Hp2, the EDRA model learns that $\mathcal{M}_{top} \gg^{tar} \mathcal{F}$ **Step 1**
  - By Hp3 and Hp4, that suffices for restrictiveness **Step 2**

EDRAs' restrictiveness on $\mathcal{F}$-irrelevant languages

Third result of this talk

The EDRA model is restrictive provided: [Magri in preparation]

Hp1 the target/training phonotactic pattern is $\mathcal{F}$-irrelevant

Hp2 $p \leq \frac{1}{w + \Delta}$ with $\Delta(m) \simeq \frac{m}{k \log m}$

Hp3 the candidacy relation is symmetric

Hp4 faithfulness constraints are distinctive

- Idea of the proof:
  - By Hp1, the target ranking is $\mathcal{M}_{top} \gg \mathcal{F} \gg \mathcal{M}_{bottom}$
  - By Hp2, the EDRA model learns that $\mathcal{M}_{top} \gg^{tar} \mathcal{F}$ **Step 1**
  - By Hp3 and Hp4, that suffices for restrictiveness **Step 2**

Step 1: EDRAs learn $\mathcal{M}_{top} \gg \mathcal{F}$

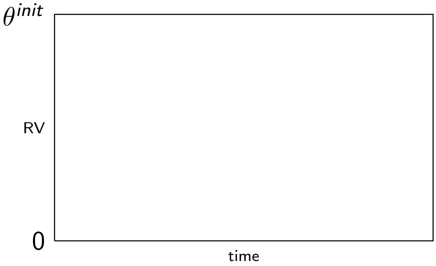
- EDRA only demotes as much as needed: [Tesar and Smolensky 1998]
  - the top ranked constraint is never demoted
  - the second top ranked constraint demoted by at most 1
  - the third top ranked constraint demoted by at most 2
  - ...

- The EDRA's ranking dynamics for $\mathcal{M}_{top}$ thus looks as follows:

Step 1: EDRAs learn $\mathcal{M}_{top} \gg \mathcal{F}$

- EDRA only demotes as much as needed: [Tesar and Smolensky 1998]
  - the top ranked constraint is never demoted
  - the second top ranked constraint demoted by at most 1
  - the third top ranked constraint demoted by at most 2
  - ...

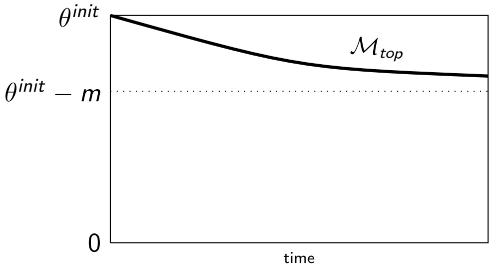
- The EDRA's ranking dynamics for $\mathcal{M}_{top}$ thus looks as follows:



Step 1: EDRAs learn $\mathcal{M}_{top} \gg \mathcal{F}$

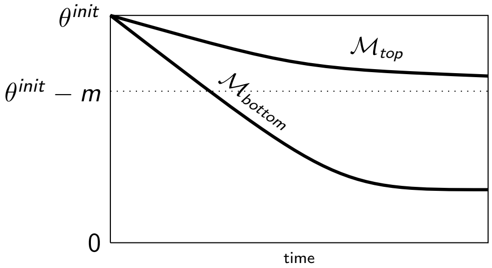
- EDRA only demotes as much as needed: [Tesar and Smolensky 1998]
  - the top ranked constraint is never demoted
  - the second top ranked constraint demoted by at most 1
  - the third top ranked constraint demoted by at most 2
  - ...

- The EDRA's ranking dynamics for $\mathcal{M}_{top}$ thus looks as follows:



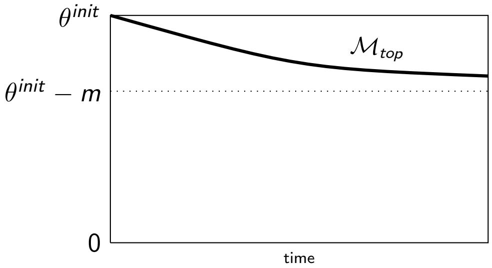
Step 1: EDRAs learn $\mathcal{M}_{top} \gg \mathcal{F}$

- EDRA only demotes as much as needed: [Tesar and Smolensky 1998]
  - the top ranked constraint is never demoted
  - the second top ranked constraint demoted by at most 1
  - the third top ranked constraint demoted by at most 2
  - ...
- The EDRA's ranking dynamics for $\mathcal{M}_{top}$ thus looks as follows:



Step 1: EDRAs learn $\mathcal{M}_{top} \gg \mathcal{F}$

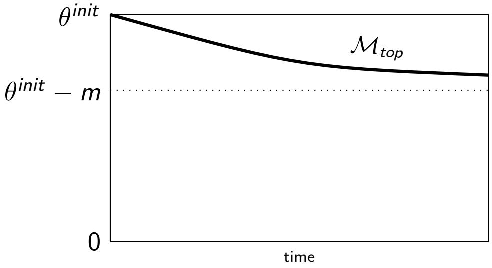
- EDRA only demotes as much as needed: [Tesar and Smolensky 1998]
  - the top ranked constraint is never demoted
  - the second top ranked constraint demoted by at most 1
  - the third top ranked constraint demoted by at most 2
  - ...
- The EDRA's ranking dynamics for $\mathcal{M}_{top}$ thus looks as follows:



- $\mathcal{M}_{top}$ need to be ranked high and indeed stay high

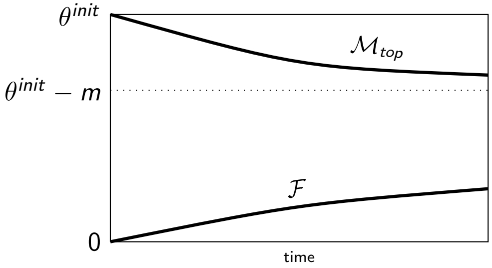
Step 1: EDRAs learn $\mathcal{M}_{top} \gg \mathcal{F}$

- As we have seen, $\mathcal{M}_{top}$ stay above $\theta^{init} - m$:



Step 1: EDRAs learn $\mathcal{M}_{top} \gg \mathcal{F}$

- As we have seen, $\mathcal{M}_{top}$ stay above $\theta^{init} - m$:



- If promotion amount p is null or small, $\mathcal{F}$ stays lower than $\theta^{init} - m$:

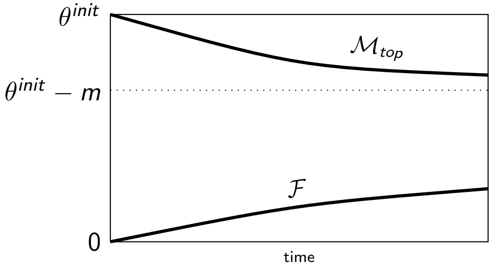
$$p = \frac{1}{w + \Delta} \qquad \Delta(m) \simeq \frac{m}{k \log m}$$

- As good as we might have hoped for:
  - $\Delta(m)$ needs to increase with m and cannot be constant
  - yet, the rate of increase (derivative) of $\Delta(m)$ decreases with m

- The target ranking condition $\mathcal{M}_{top} \gg \mathcal{F}$ is thus learned

Step 1: EDRAs learn $\mathcal{M}_{top} \gg \mathcal{F}$

- As we have seen, $\mathcal{M}_{top}$ stay above $\theta^{init} - m$:



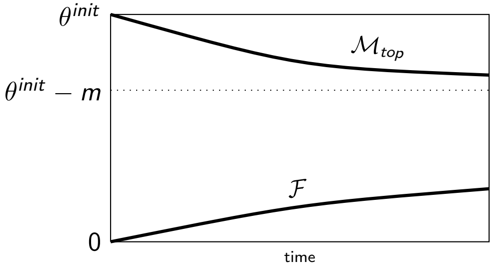
- If promotion amount p is null or small, $\mathcal{F}$ stays lower than $\theta^{init} - m$:

$$p = \frac{1}{w + \Delta} \qquad \Delta(m) \simeq \frac{m}{k \log m}$$

- As good as we might have hoped for:
  - $\Delta(m)$ needs to increase with m and cannot be constant
  - yet, the rate of increase (derivative) of $\Delta(m)$ decreases with m
- The target ranking condition $\mathcal{M}_{top} \gg \mathcal{F}$ is thus learned

Step 1: EDRAs learn $\mathcal{M}_{top} \gg \mathcal{F}$

- As we have seen, $\mathcal{M}_{top}$ stay above $\theta^{init} - m$:



- If promotion amount p is null or small, $\mathcal{F}$ stays lower than $\theta^{init} - m$:

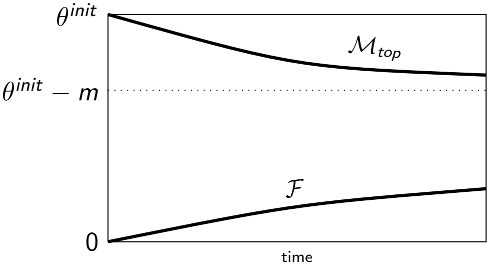
$$p = \frac{1}{w + \Delta} \qquad \Delta(m) \simeq \frac{m}{k \log m}$$

- As good as we might have hoped for:
  - $\Delta(m)$ needs to increase with m and cannot be constant
  - yet, the rate of increase (derivative) of $\Delta(m)$ decreases with m

- The target ranking condition $\mathcal{M}_{top} \gg \mathcal{F}$ is thus learned

Step 1: EDRAs learn $\mathcal{M}_{top} \gg \mathcal{F}$

- As we have seen, $\mathcal{M}_{top}$ stay above $\theta^{init} - m$:



- If promotion amount p is null or small, $\mathcal{F}$ stays lower than $\theta^{init} - m$:

$$p = \frac{1}{w + \Delta} \qquad \Delta(m) \simeq \frac{m}{k \log m}$$

- As good as we might have hoped for:
  - $\Delta(m)$ needs to increase with m and cannot be constant
  - yet, the rate of increase (derivative) of $\Delta(m)$ decreases with m

- The target ranking condition $\mathcal{M}_{top} \gg \mathcal{F}$ is thus learned

Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology

target ranking	**prefers** /**x**/→[**x**] vs	/**x**/→[**y**]
M		$\surd$
$M' M''\ldots$	$\surd$	
$F' F''\ldots$	$\surd$	

Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology

target ranking	**learned ranking**	**prefers** /**x**/→[**x**]	vs /**x**/→[**y**]
M			$\surd$
$M' M'' \ldots$		$\surd$	
$F' F'' \ldots$		$\surd$	

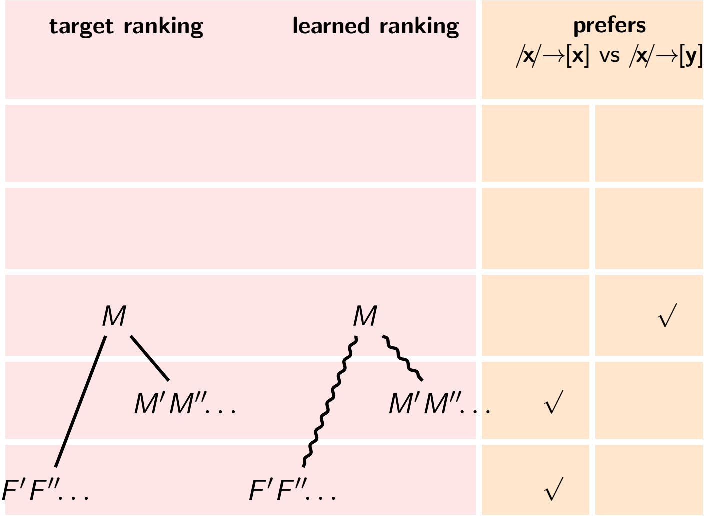
Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology

target ranking	**learned ranking**	**prefers** /**x**/→[**x**]	vs /**x**/→[**y**]
M	M		√
$M' M'' \ldots$		√	
$F' F'' \ldots$	$F' F'' \ldots$	√	

Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology



Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology

target ranking	**learned ranking**	**prefers** /**x**/→[**x**]	vs /**x**/→[**y**]
	M'	√	
M	M		√
$M' M'' \ldots$	$M'' \ldots$	√	
$F' F'' \ldots$	$F' F'' \ldots$	√	

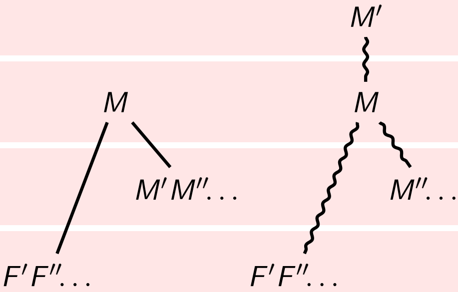
Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology

target ranking	**learned ranking**	**prefers** /**x**/→[**x**] vs /**x**/→[**y**]		**prefers** /**y**/→[**x**] vs /**y**/→[**y**]	
	M'	$\surd$			
M	M		$\surd$		
$M'\,M''\ldots$	$M''\ldots$	$\surd$			
$F'\,F''\ldots$	$F'\,F''\ldots$	$\surd$			

Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology



target ranking	learned ranking	prefers /**x**/→[**x**] vs /**x**/→[**y**]		prefers /**y**/→[**x**] vs /**y**/→[**y**]	
	M'	√		√	
M	M		√		
$M' M'' \ldots$	$M'' \ldots$	√			
$F' F'' \ldots$	$F' F'' \ldots$	√			

Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology

target ranking	**learned ranking**	**prefers** /**x**/→[**x**] vs /*x*/→[**y**]		**prefers** /**y**/→[**x**] vs /**y**/→[**y**]	
	C				√
	M'	√		√	
M	M		√		
$M' M'' \ldots$	$M'' \ldots$	√			
$F' F'' \ldots$	$F' F'' \ldots$	√			

Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology

target ranking	**learned ranking**	**prefers** /**x**/→[**x**]	vs /**x**/→[**y**]	**prefers** /**y**/→[**x**]	vs /**y**/→[**y**]
	F				√
	M'	√		√	
M	M		√		
$M'\,M''\ldots$	$M''\ldots$	√			
$F'\,F''\ldots$	$F'\,F''\ldots$	√			

Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology

target ranking	**learned ranking**	**prefers** /**x**/→[**x**] vs /**x**/→[**y**]		**prefers** /**y**/→[**x**] vs /**y**/→[**y**]	
	C				√
	M'	√		√	
M	M		√		
$M' M'' \dots$	$M'' \dots$	√			
$F' F'' \dots$	$F' F'' \dots$	√			

Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology

target ranking	**learned ranking**	**prefers** /**x**/→[**x**] vs /**x**/→[**y**]		**prefers** /**y**/→[**x**] vs /**y**/→[**y**]	
	$\overline{M}$				√
	M'	√		√	
M	M		√		
$M' M'' \ldots$	$M'' \ldots$	√			
$F' F'' \ldots$	$F' F'' \ldots$	√			

Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology

target ranking	**learned ranking**	**prefers** /**x**/→[x] vs /**x**/→[**y**]		**prefers** /**y**/→[**x**] vs /**y**/→[**y**]	
	$\overline{M}$		√		√
	M'	√		√	
M	M		√		
$M' M'' \ldots$	$M'' \ldots$	√			
$F' F'' \ldots$	$F' F'' \ldots$	√			

Step 2: $\mathcal{M}_{top} \gg \mathcal{F}$ suffices for restrictiveness

Suppose **x** is illicit and thus neutralized to **y** by the target phonology

target ranking	learned ranking	prefers /**x**/→[**x**]	vs /**x**/→[**y**]
	$\overline{M}$		√
	M'	√	
M			√
$M' M'' \dots$	$M'' \dots$	√	
$F' F'' \dots$	$F' F'' \dots$	√	

5.
Conclusions

Concluding remark I

Second result of this talk

The OT problem of the acquisition of phonotactics (stated as a Subset problem) is intractable, even in the best conditions [Magri 2013]

- GL usually motivated through poverty of the stimulus arguments: child's linguistic input is ambiguous, incomplete, degenerate [Thomas 2002]
- Poverty of the stimulus arguments are difficult to make:
  - empirical side: they are about child input
  - theoretical side: what suffices for learnability [Clark and Lappin 2011]
- Results such as the one above show that learning is hard even when the input is rich and idealized:
  - pairs of underlying and surface representations
  - pristine and uncorrupted
  - no hidden structure
- From poverty of the stimulus to hardness of the task: intractability results provide further ammunition for GL

Concluding remark I

Second result of this talk

The OT problem of the acquisition of phonotactics (stated as a Subset problem) is intractable, even in the best conditions [Magri 2013]

- GL usually motivated through poverty of the stimulus arguments: child's linguistic input is ambiguous, incomplete, degenerate [Thomas 2002]
- Poverty of the stimulus arguments are difficult to make:
  - empirical side: they are about child input
  - theoretical side: what suffices for learnability [Clark and Lappin 2011]
- Results such as the one above show that learning is hard even when the input is rich and idealized:
  - pairs of underlying and surface representations
  - pristine and uncorrupted
  - no hidden structure
- From poverty of the stimulus to hardness of the task: intractability results provide further ammunition for GL

Concluding remark I

Second result of this talk

The OT problem of the acquisition of phonotactics (stated as a Subset problem) is intractable, even in the best conditions [Magri 2013]

- GL usually motivated through poverty of the stimulus arguments: child's linguistic input is ambiguous, incomplete, degenerate [Thomas 2002]
- Poverty of the stimulus arguments are difficult to make:
  - empirical side: they are about child input
  - theoretical side: what suffices for learnability [Clark and Lappin 2011]
- Results such as the one above show that learning is hard even when the input is rich and idealized:
  - pairs of underlying and surface representations
  - pristine and uncorrupted
  - no hidden structure
- From poverty of the stimulus to hardness of the task: intractability results provide further ammunition for GL

Concluding remark I

Second result of this talk

The OT problem of the acquisition of phonotactics (stated as a Subset problem) is intractable, even in the best conditions [Magri 2013]

- GL usually motivated through poverty of the stimulus arguments: child's linguistic input is ambiguous, incomplete, degenerate [Thomas 2002]
- Poverty of the stimulus arguments are difficult to make:
  - empirical side: they are about child input
  - theoretical side: what suffices for learnability [Clark and Lappin 2011]
- Results such as the one above show that learning is hard even when the input is rich and idealized:
  - pairs of underlying and surface representations
  - pristine and uncorrupted
  - no hidden structure
- From poverty of the stimulus to hardness of the task: intractability results provide further ammunition for GL

Concluding remark I

Second result of this talk

The OT problem of the acquisition of phonotactics (stated as a Subset problem) is intractable, even in the best conditions [Magri 2013]

- GL usually motivated through poverty of the stimulus arguments: child's linguistic input is ambiguous, incomplete, degenerate [Thomas 2002]
- Poverty of the stimulus arguments are difficult to make:
  - empirical side: they are about child input
  - theoretical side: what suffices for learnability [Clark and Lappin 2011]
- Results such as the one above show that learning is hard even when the input is rich and idealized:
  - pairs of underlying and surface representations
  - pristine and uncorrupted
  - no hidden structure
- From poverty of the stimulus to hardness of the task: intractability results provide further ammunition for GL

Concluding remark II

First result of this talk

The EDRA model converges efficiently under the OT mode of constraint interaction ("tilted planes" assumption) [Tesar and Smolensky 1998; Pater 2008; Magri 2012]

Third result of this talk

The EDRA model is restrictive when the target/training phonotactic pattern is $\mathcal{F}$-irrelevant (under mild additional assumptions) [Magri in preparation]

- Current skepticism about phonological universals: [Evans and Levinson 2009]
  - "the study of universals is fraught with difficulties" [Hyman 2008]
  - "there appear to be so few absolute universals" [Maddieson 1984]
  - "the quest for universals has failed" [van Oostendorp 2013]
- This is because phonology has focused on the wrong type of universals: "every language has a coronal stop" [Hyman 2008; Blevins 2009]

Concluding remark II

First result of this talk

The EDRA model converges efficiently under the OT mode of constraint interaction ("tilted planes" assumption) [Tesar and Smolensky 1998; Pater 2008; Magri 2012]

Third result of this talk

The EDRA model is restrictive when the target/training phonotactic pattern is $\mathcal{F}$-irrelevant (under mild additional assumptions) [Magri in preparation]

- Current skepticism about phonological universals: [Evans and Levinson 2009]
  - "the study of universals is fraught with difficulties" [Hyman 2008]
  - "there appear to be so few absolute universals" [Maddieson 1984]
  - "the quest for universals has failed" [van Oostendorp 2013]
- This is because phonology has focused on the wrong type of universals: "every language has a coronal stop" [Hyman 2008; Blevins 2009]

Concluding remark II

First result of this talk

The EDRA model converges efficiently under the OT mode of constraint interaction ("tilted planes" assumption) [Tesar and Smolensky 1998; Pater 2008; Magri 2012]

Third result of this talk

The EDRA model is restrictive when the target/training phonotactic pattern is $\mathcal{F}$-irrelevant (under mild additional assumptions) [Magri in preparation]

- Current skepticism about phonological universals: [Evans and Levinson 2009]
  - "the study of universals is fraught with difficulties" [Hyman 2008]
  - "there appear to be so few absolute universals" [Maddieson 1984]
  - "the quest for universals has failed" [van Oostendorp 2013]
- This is because phonology has focused on the wrong universals: "every language has a coronal stop" [Hyman 2008; Blevins 2009]
- The right universals are motivated by computational considerations (learnability, algorithm for production and interpretation)

Thanks!

References I

Barton, G. Edward, Robert Berwick, and Eric Sven Ristad. 1987. *Computational complexity and natural language*. Cambridge, MA: MIT Press.

Beckman, Mary E. 1999. *Positional faithfulness: An Oprimality Theretic treatment of phonological asymmetries*. New York: Garland Publishing (Outstanding dissertations in Linguistics).

Blevins, Juliette. 2009. Another universal bites the dust: Northwest Mekeo lacks coronal phonemes. *Oceanic Linguistics* 48.1:264–273.

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.

Cesa-Bianchi, Nicolò, and Gábor Lugosi. 2006. *Prediction, learning, and games*. Cambridge University Press.

Clark, Alexander, and Shalom Lappin. 2011. *Linguistic nativism and the poverty of the stimulus*. Wiley-Blackwell.

References II

Davidson, Lisa, Peter W. Jusczyk, and Paul Smolensky. 2004. The initial and final states: Theoretical implications and experimental explorations of richness of the base. In *Constraints in phonological acquisition*, ed. R. Kager, J. Pater, and W. Zonneveld, 158–203. Cambridge University Press.

Evans, Nicholas, and Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32.5:429–448.

Fodor, Janet Dean, and William Gregory Sakas. 2005. The Subset Principle in syntax: costs of compliance. *Linguistics* 41:513–569.

Galil, Zvi, and Nimrod Megiddo. 1977. Cyclic ordering is NP-complete. *Theoretical Computer Science* 5:179–182.

Gibson, Edward, and Kenneth Wexler. 1994. Triggers. *Linguistic Inquiry* 25:407–454.

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.

References III

Hayes, Bruce, and Donca Steriade. 2004. Introduction: the phonetic bases of phonological markedness. In *Phonetically based phonology*, ed. Bruce Hayes, Robert Kirchner, and Donca Steriade, 1–33. Cambridge University Press.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.

Hyman, Larry. 2008. Universals in phonology. In *The linguistic review*, ed. Harry van der Hulst.

Idsardi, William. 2006. A simple proof that Optimality Theory is computationally intractable. *Linguistic Inquiry* 37.2:271–275.

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.

Jusczyk, Peter, Paul Smolensky, and Theresa Allocco. 2002. How english-learning infants respond to markedness and faithfulness constraints. *Language Acquisition* 10:31–73.

Kazazis, Kostas. 1969. Possible evidence for (near-)underlying forms in the speech of a child. In *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society (CLS5)*, ed. Robert I. Binnick, Alice Davison, Georgia McGreen, and Jerry L. Morgan, 382–388. Chicago, IL: Chicago Linguistic Society.

References IV

Kenstowicz, Michael, and Charles W. Kisseberth. 1977. *Topics in phonological theory*. New York: Academic Press.

Legendre, Gèraldine, Antonella Sorace, and Paul Smolensky. 2006. The optimality theory/harmonic grammar connection. In *The harminic mind*, ed. Paul Smolensky and Gèraldine Legendre, 903–966. Cambridge, MA: MIT Press.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge University Press.

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

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

Magri, Giorgio. 2016. Idempotency in Optimality Theory. Submitted manuscript.

Magri, Giorgio. in preparation. Some restrictiveness results for the calibrated error-driven ranking algorithm. Manuscript.

Magri, Giorgio. to appeara. Error-driven learning in OT and HG: a comparison. *Phonology* .

Magri, Giorgio. to appearb. Noise robustness and stochastic tolerance of OT error-driven ranking algorithms. *Journal of Logic and Computation* .

References V

Magri, Giorgio, and René Kager. 2015. How to choose successful losers. In *Proceedings of MOL 14: the 14th meeting on Mathematics of Language*.

Mazuka, Reiko, Yvonne Cao, Emmanuel Dupoux, and Anne Christophe. 2011. The development of a phonological illusion: a cross-linguistic study with japanese and french infants. *Developmental Science* 14.4:693–699.

McLeod, S., J. van Doorn, and V. Reed. 2001. Consonant cluster development in two-year-olds: General trends and individual differences. *Journal of Speech, Language and Hearing Research* 44:1144–1172.

Moreton, Elliott. 2004. Non-computable functions in Optimality Theory. In *Optimality theory in phonology: A reader*, ed. John J. McCarthy, 141–163. Malden: MA: Wiley-Blackwell.

van Oostendorp, Marc. 2013. Phonology between theory and data. In *L'interface langage-cognition / the language-cognition interface. actes du 19e congrès international des linguistes (langue et cultures, 45)*, ed. S.R. Anderson, J. Moeschler, and F. Reboul, 289–306. Genève - Paris: Libraire Droz.

Pater, Joe. 2007. "the power of weighted constraints". Talk delivered at Johns Hopkins.

Pater, Joe. 2008. Gradual learning and convergence. *Linguistic Inquiry* 39:334–345.

Pinker, S. 1979. Formal models of language learning. *Cognition* 7:217–282.

References VI

Prince, Alan, and Paul Smolensky. 1997. "optimality: From neural networks to universal grammar". *Science* 275:1604–1610.

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.

Prince, Alan, and Bruce Tesar. 2008. Using phonotactics to learn phonological alternations. In *Proceedings of CLS 39*, –.

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

Smolensky, Paul. 1996b. On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27.4:720–731.

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

Thomas, M. 2002. Development of the concept of the poverty of the stimulus. *The Linguistic Review* 18.1-2:51–71.