T-orders across categorical and probabilistic constraint-based phonology

Arto Anttila (Stanford University) and Giorgio Magri (CNRS)

AMP 2017, New York University, September 15-17

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR
ANR

FRANCE-STANFORD CENTER FOR INTERDISCIPLINARY STUDIES

Introduction

- **Implicational universals**: [Greenberg 1963]
  - $P \longrightarrow \widehat{P}$
    if a language has property P, it also has property $\widehat{P}$

- **Implicational universals**: [Greenberg 1963]
  - $P \longrightarrow \widehat{P}$
    if a language has property P, it also has property $\widehat{P}$

- **Examples:**
  - $CCCV \longrightarrow CCV \longrightarrow CV$
    if a language allows a complex margin cluster, it allows a simpler one
  - $(/cost.us/, [cos.us]) \longrightarrow (/cost.me/, [cos.me])$
    if a dialect of English deletes t/d before V, it also deletes it before C

- Consider a typology $\mathfrak{T}$ of phonological grammars, construed as mappings from underlying representations to surface representations

- Consider a typology $\mathfrak{T}$ of phonological grammars, construed as mappings from underlying representations to surface representations
- Consider the simplest antecedent and consequent properties P and $\widehat{P}$:
  - P: the property of mapping an ur x to a sr y: (x, y)
  - $\widehat{P}$: the property of mapping an ur $\widehat{x}$ to a sr $\widehat{y}$: $(\widehat{x}, \widehat{y})$

- Consider a typology $\mathfrak{T}$ of phonological grammars, construed as mappings from underlying representations to surface representations
- Consider the simplest antecedent and consequent properties P and $\widehat{P}$:
  - P: the property of mapping an ur x to a sr y: (x, y)
  - $\widehat{P}$: the property of mapping an ur $\widehat{x}$ to a sr $\widehat{y}$: $(\widehat{x}, \widehat{y})$
- The implicational universal

$$(x, y) \xrightarrow{\mathfrak{T}} (\widehat{x}, \widehat{y})$$

holds provided each grammar in $\mathfrak{T}$ which maps x to y also maps $\widehat{x}$ to $\widehat{y}$

- Consider a typology $\mathfrak{T}$ of phonological grammars, construed as mappings from underlying representations to surface representations
- Consider the simplest antecedent and consequent properties P and $\widehat{P}$:
  - P: the property of mapping an ur x to a sr y: (x, y)
  - $\widehat{P}$: the property of mapping an ur $\widehat{x}$ to a sr $\widehat{y}$: $(\widehat{x}, \widehat{y})$
- The implicational universal

$$(x, y) \xrightarrow{\mathfrak{T}} (\widehat{x}, \widehat{y})$$

  holds provided each grammar in $\mathfrak{T}$ which maps x to y also maps $\widehat{x}$ to $\widehat{y}$
- $\xrightarrow{\mathfrak{T}}$ is a partial order called the **T-order** induced by $\mathfrak{T}$ [Anttila and Andrus 2006]

- Implicational universals can be statistical: [Guy 1991; Kiparsky 1993; Coetzee 2004]
  - variable t/d deletion is more frequent before C than V

- Implicational universals can be statistical: [Guy 1991; Kiparsky 1993; Coetzee 2004]
  - variable t/d deletion is more frequent before C than V

- Consider a typology $\mathfrak{T}$ of **probabilistic** phonological grammars, construed as functions from underlying forms to probability distributions over surface forms

- Implicational universals can be statistical: [Guy 1991; Kiparsky 1993; Coetzee 2004]
  - variable t/d deletion is more frequent before C than V
- Consider a typology $\mathfrak{T}$ of **probabilistic** phonological grammars, construed as functions from underlying forms to probability distributions over surface forms
- The implicational universal

$$(x, y) \xrightarrow{\mathfrak{T}} (\widehat{x}, \widehat{y})$$

holds provided each grammar in $\mathfrak{T}$ assigns a probability to $(\widehat{x}, \widehat{y})$ which is at least as large as the probability it assigns to (x, y)

- Implicational universals can be statistical: [Guy 1991; Kiparsky 1993; Coetzee 2004]
  - variable t/d deletion is more frequent before C than V
- Consider a typology $\mathfrak{T}$ of **probabilistic** phonological grammars, construed as functions from underlying forms to probability distributions over surface forms
- The implicational universal

$$(x, y) \xrightarrow{\mathfrak{T}} (\widehat{x}, \widehat{y})$$

holds provided each grammar in $\mathfrak{T}$ assigns a probability to $(\widehat{x}, \widehat{y})$ which is at least as large as the probability it assigns to (x, y)

- The preceding categorical definition of T-orders is a special case of the probabilistic one

- T-orders are interesting/important because:
  - impose strict limits on categorical and statistical phonological patterns
  - "measure" the amount of typological structure
  - follow from the phonological theory
  - need not be learned, cannot be subverted by learning [Becker et al. 2011]
  - model categorical and gradient phonotactic judgments [Anttila 2008]

- T-orders are interesting/important because:
  - impose strict limits on categorical and statistical phonological patterns
  - "measure" the amount of typological structure
  - follow from the phonological theory
  - need not be learned, cannot be subverted by learning [Becker et al. 2011]
  - model categorical and gradient phonotactic judgments [Anttila 2008]

- We develop formal theory of T-orders in constraint-based phonology:
  - two categorical frameworks: categorical OT ($\xrightarrow{\text{OT}}$)
    categorical HG ($\xrightarrow{\text{HG}}$)
  - four probabilistic frameworks: stochastic OT ($\xrightarrow{\text{sOT}}$)
    stochastic (or noisy) HG ($\xrightarrow{\text{sHG}}$)
    partial order OT ($\xrightarrow{\text{poOT}}$)
    Max Ent ($\xrightarrow{\text{ME}}$)

  and we explore its phonological implications

- **Formal results 1**: a complete characterization of the relationships among T-orders in these six frameworks
  - T-orders indeed allow for cross-framework comparisons of typological structure, even bridging across categorical and probabilistic frameworks

- **Formal results 1**: a complete characterization of the relationships among T-orders in these six frameworks
  - T-orders indeed allow for cross-framework comparisons of typological structure, even bridging across categorical and probabilistic frameworks
- **Formal results 2**: an almost complete characterization of T-orders in the six frameworks in terms of just constraint violation profiles of the antecedent mapping, the consequent mapping, and their losers
  - help us understand what it really means that a T-order holds
  - allow us to compute T-orders without computing the entire typology

- **Formal results 1**: a complete characterization of the relationships among T-orders in these six frameworks
  - T-orders indeed allow for cross-framework comparisons of typological structure, even bridging across categorical and probabilistic frameworks
- **Formal results 2**: an almost complete characterization of T-orders in the six frameworks in terms of just constraint violation profiles of the antecedent mapping, the consequent mapping, and their losers
  - help us understand what it really means that a T-order holds
  - allow us to compute T-orders without computing the entire typology
- **Phonological implications 1**: a surprising equivalence result for OT and HG T-orders under the 2-candidate condition

- **Formal results 1**: a complete characterization of the relationships among T-orders in these six frameworks
  - T-orders indeed allow for cross-framework comparisons of typological structure, even bridging across categorical and probabilistic frameworks
- **Formal results 2**: an almost complete characterization of T-orders in the six frameworks in terms of just constraint violation profiles of the antecedent mapping, the consequent mapping, and their losers
  - help us understand what it really means that a T-order holds
  - allow us to compute T-orders without computing the entire typology
- **Phonological implications 1**: a surprising equivalence result for OT and HG T-orders under the 2-candidate condition
- **Phonological implications 2**: first ever analytical results on MaxEnt T-orders, showing that they display some counterintuitive properties

Formal results 1

$$\boxed{(x, y) \xrightarrow{\text{ME}} (\widehat{x}, \widehat{y})}$$

$$\boxed{(x, y) \xrightarrow{\text{poOT}} (\widehat{x}, \widehat{y})}$$

$$\boxed{(x, y) \xrightarrow{\text{HG}} (\widehat{x}, \widehat{y})}$$

$$\boxed{(x, y) \xrightarrow{\text{OT}} (\widehat{x}, \widehat{y})}$$

$$\boxed{(x, y) \xrightarrow{\text{sHG}} (\widehat{x}, \widehat{y})}$$

$$\boxed{(x, y) \xrightarrow{\text{sOT}} (\widehat{x}, \widehat{y})}$$

$$\boxed{(x, y) \xrightarrow{\text{ME}} (\widehat{x}, \widehat{y})} \qquad \boxed{(x, y) \xrightarrow{\text{poOT}} (\widehat{x}, \widehat{y})}$$

$$\boxed{(x, y) \xrightarrow{\text{HG}} (\widehat{x}, \widehat{y})} \longrightarrow \boxed{(x, y) \xrightarrow{\text{OT}} (\widehat{x}, \widehat{y})}$$

$$\boxed{(x, y) \xrightarrow{\text{sHG}} (\widehat{x}, \widehat{y})} \qquad \boxed{(x, y) \xrightarrow{\text{sOT}} (\widehat{x}, \widehat{y})}$$

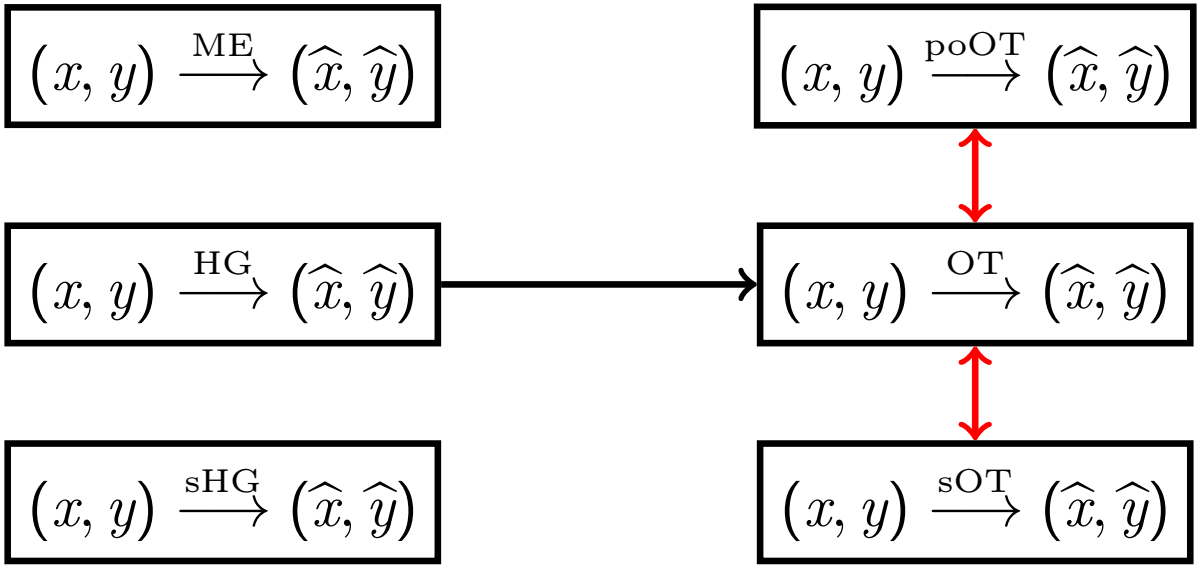
- The HG typology is usually at least as large as the OT typology
  hence, a HG T-order entails the OT T-order
  and the vice versa fails in the general case

$$\boxed{(x, y) \xrightarrow{\text{ME}} (\widehat{x}, \widehat{y})} \qquad \boxed{(x, y) \xrightarrow{\text{poOT}} (\widehat{x}, \widehat{y})}$$

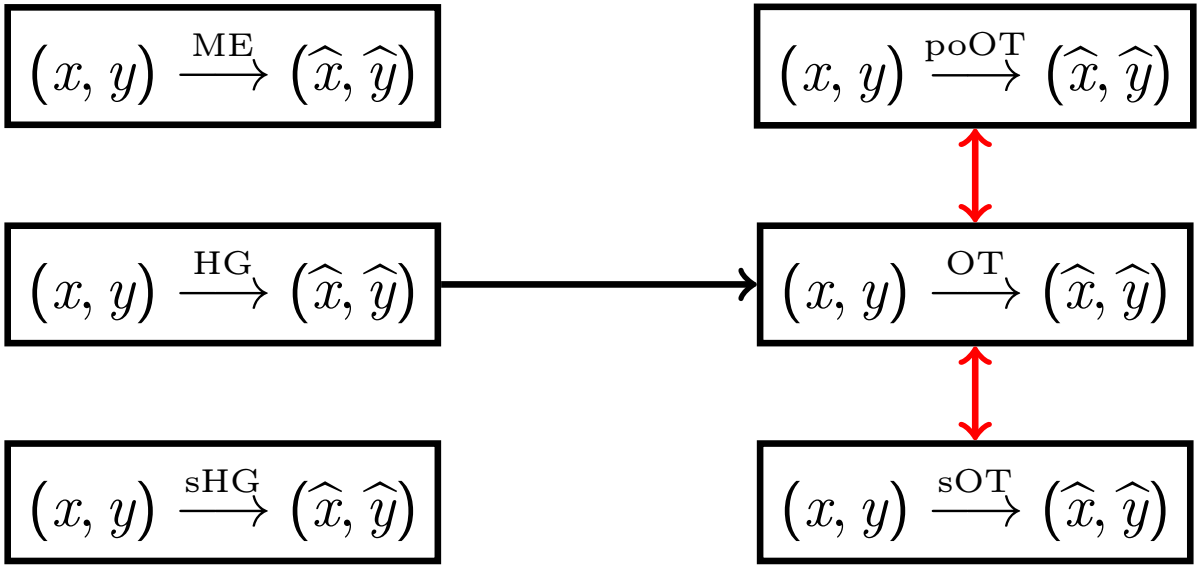
$$\boxed{(x, y) \xrightarrow{\text{HG}} (\widehat{x}, \widehat{y})} \longrightarrow \boxed{(x, y) \xrightarrow{\text{OT}} (\widehat{x}, \widehat{y})}$$

$$\boxed{(x, y) \xrightarrow{\text{sHG}} (\widehat{x}, \widehat{y})} \qquad \boxed{(x, y) \xrightarrow{\text{sOT}} (\widehat{x}, \widehat{y})}$$

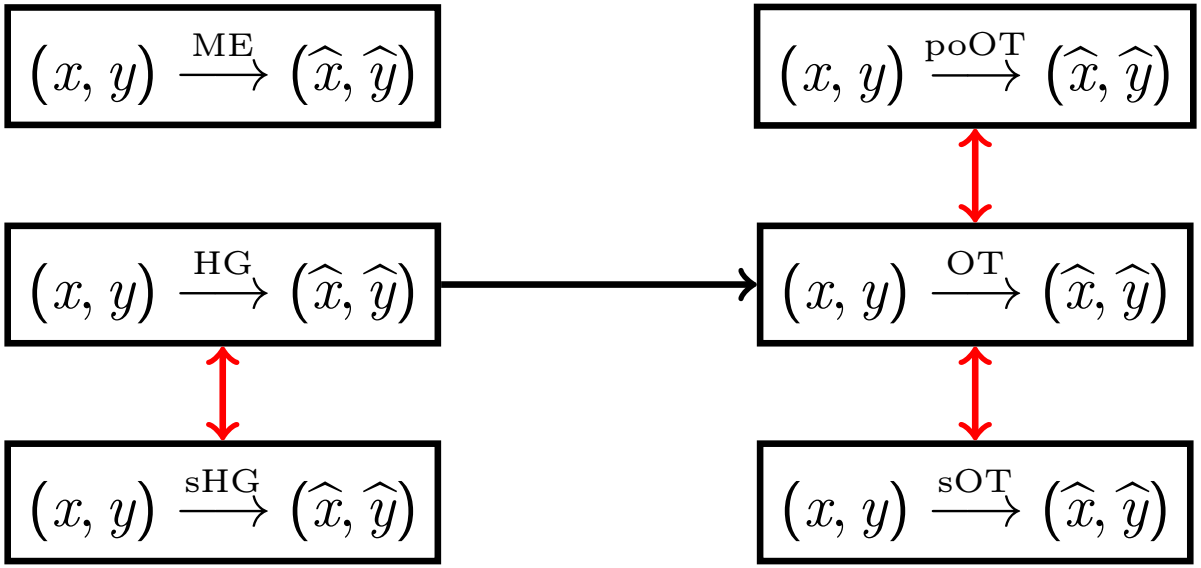
- The HG typology is usually at least as large as the OT typology hence, a HG T-order entails the OT T-order and the vice versa fails in the general case
- But see below for more on this!



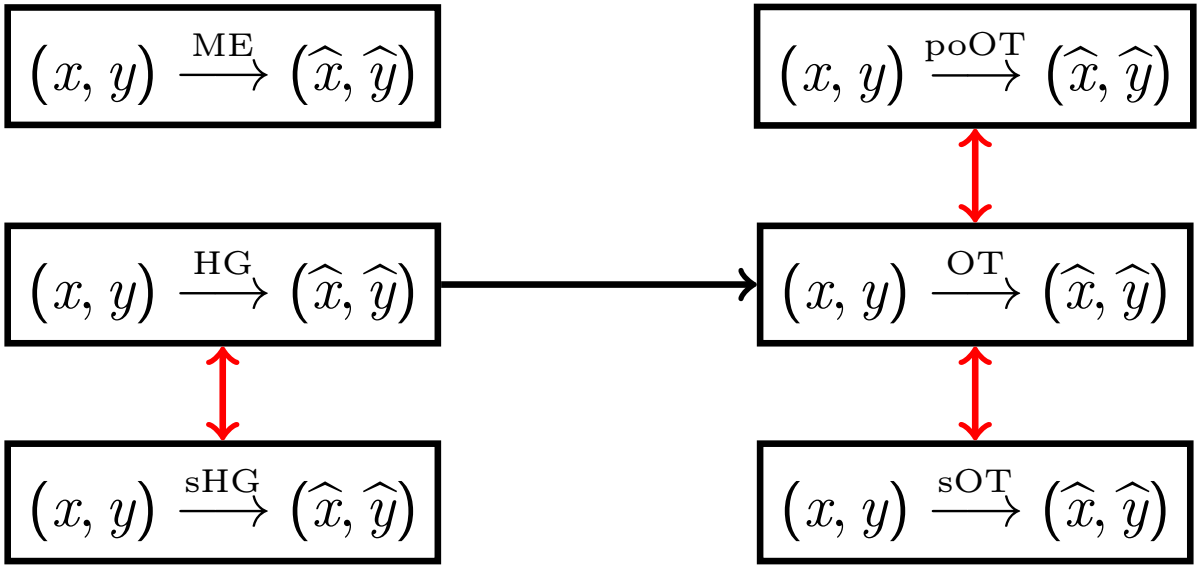
- T-orders in OT, stochastic OT, and partial order OT are identical



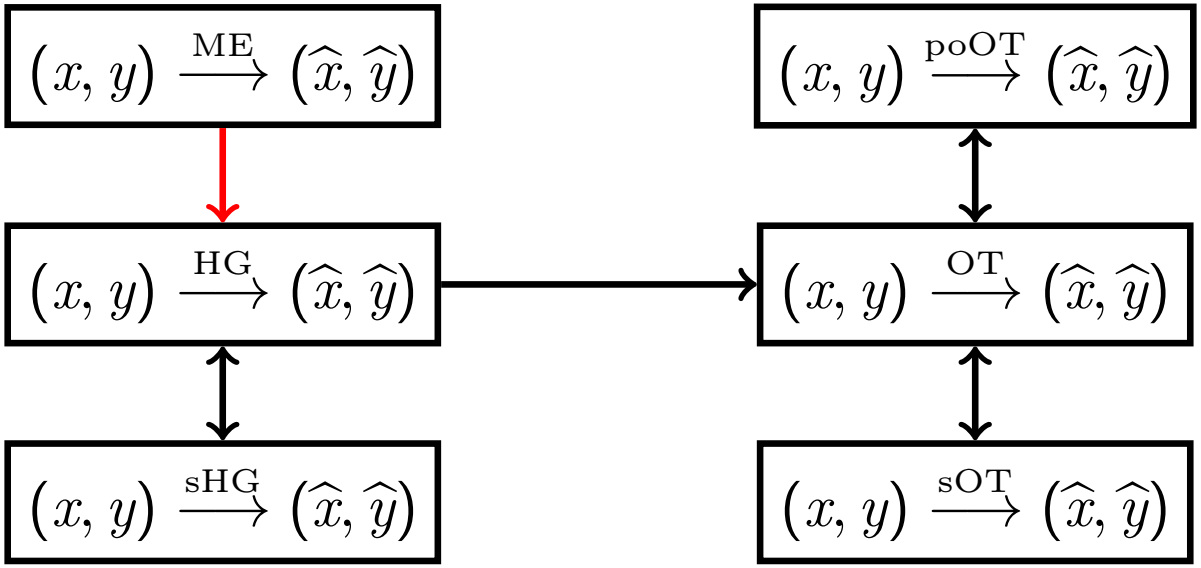
- T-orders in OT, stochastic OT, and partial order OT are identical
- despite the fact that the OT typology is finite
  the partial order OT typology is much larger
  the stochastic OT typology is infinite



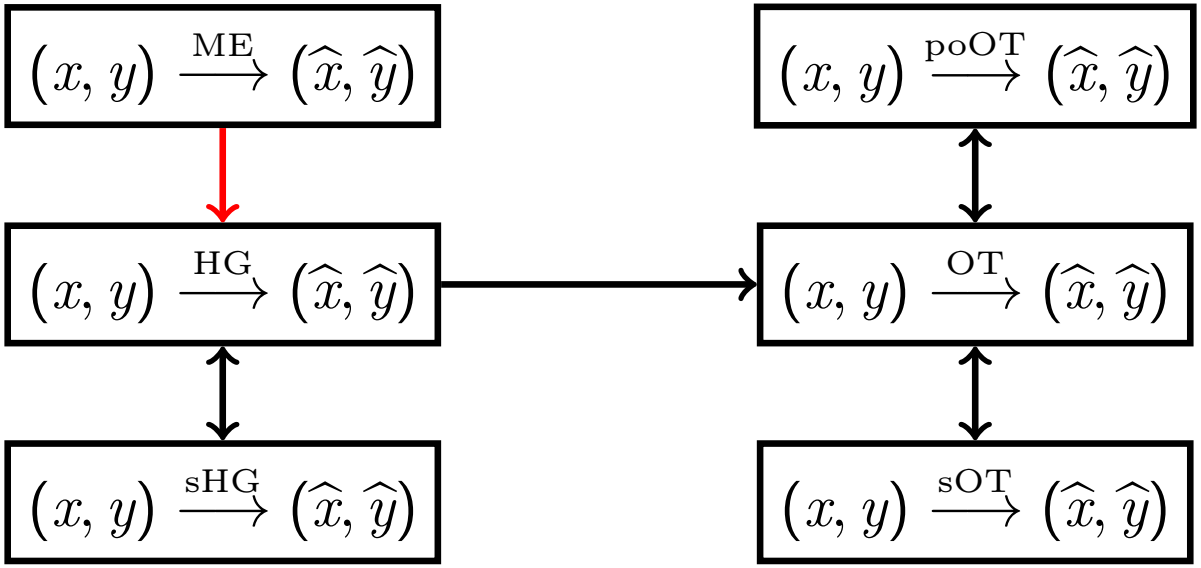
- T-orders in OT, stochastic OT, and partial order OT are identical
- despite the fact that the OT typology is finite
  the partial order OT typology is much larger
  the stochastic OT typology is infinite
- Analogous equivalence holds for HG and stochastic HG



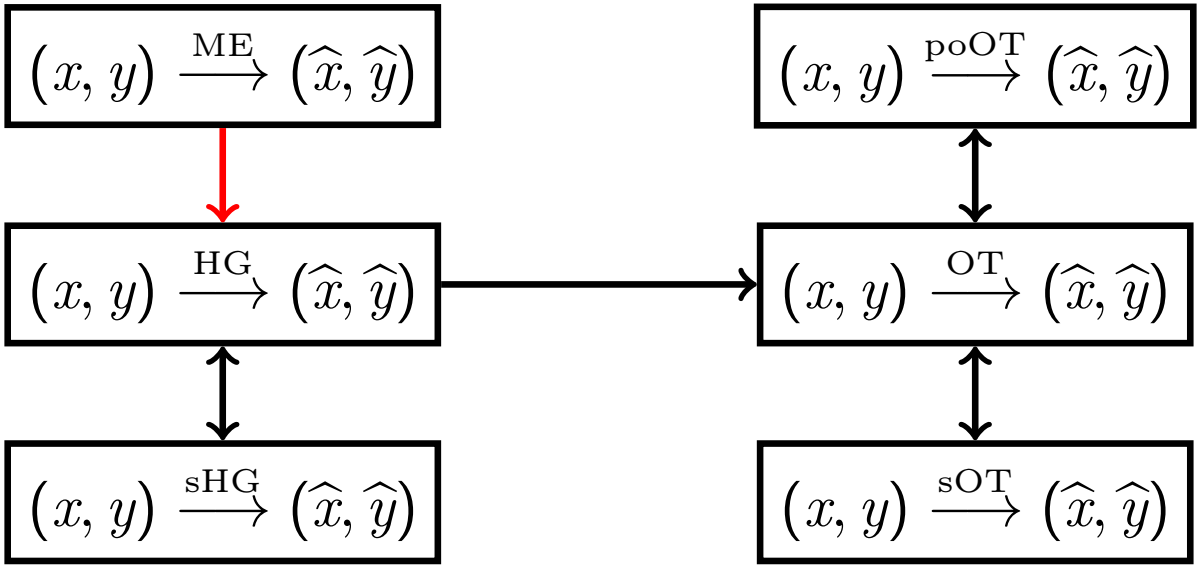
- T-orders in OT, stochastic OT, and partial order OT are identical
- despite the fact that the OT typology is finite
  the partial order OT typology is much larger
  the stochastic OT typology is infinite
- Analogous equivalence holds for HG and stochastic HG
- Thus, stochastic OT/HG are probabilistic variants of categorical OT/HG which don't tamper with the categorical typological structure



- A ME T-order entails an HG T-order



- A ME T-order entails an HG T-order
- The vice versa fails even with 1 constraint and 1 loser!



- A ME T-order entails an HG T-order
- The vice versa fails even with 1 constraint and 1 loser!
- Thus, although ME and stochastic HG both look like "innocuous" probabilistic variants of HG, ME actually loses much of the typological structure imposed by HG

Formal results 2

- Because of the equivalences above, we only need constraint conditions for a T-order in HG, OT, and MaxEnt

- Because of the equivalences above, we only need constraint conditions for a T-order in HG, OT, and MaxEnt
- Focus on the antecedent mapping (x, y) of the T-order $(x, y) \rightarrow (\widehat{x}, \widehat{y})$

- Because of the equivalences above, we only need constraint conditions for a T-order in HG, OT, and MaxEnt
- Focus on the antecedent mapping (x, y) of the T-order $(x, y) \rightarrow (\widehat{x}, \widehat{y})$
  consider a corresponding loser mapping (x, z)

- Because of the equivalences above, we only need constraint conditions for a T-order in HG, OT, and MaxEnt

- Focus on the antecedent mapping (x, y) of the T-order $(x, y) \rightarrow (\widehat{x}, \widehat{y})$
  consider a corresponding loser mapping (x, z)
  compute the **antecedent difference vector**:

$$\begin{array}{c}\text{violations of}\\ \text{the loser } (x,z)\end{array} - \begin{array}{c}\text{violations of the}\\ \text{antecedent } (x,y)\end{array} = \begin{bmatrix} C_1(x,z) - C_1(x,y) \\ C_2(x,z) - C_2(x,y) \\ \vdots \\ C_n(x,z) - C_n(x,y) \end{bmatrix}$$

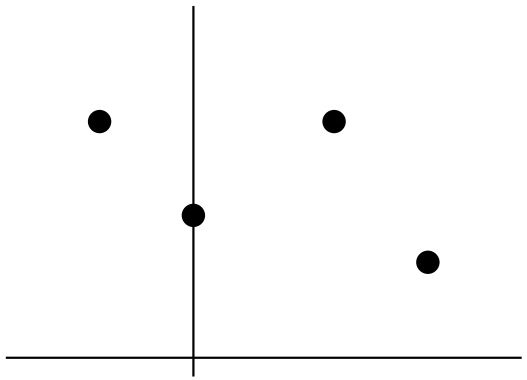
- Because of the equivalences above, we only need constraint conditions for a T-order in HG, OT, and MaxEnt

- Focus on the antecedent mapping (x, y) of the T-order $(x, y) \rightarrow (\widehat{x}, \widehat{y})$
  consider a corresponding loser mapping (x, z)
  compute the **antecedent difference vector**:

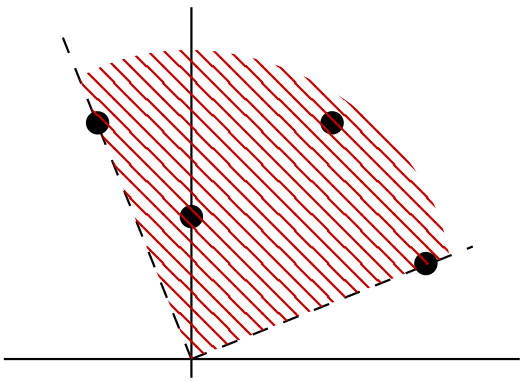
$$\begin{array}{c}\text{violations of}\\ \text{the loser } (x,z)\end{array} - \begin{array}{c}\text{violations of the}\\ \text{antecedent } (x,y)\end{array} = \begin{bmatrix} C_1(x,z) - C_1(x,y) \\ C_2(x,z) - C_2(x,y) \\ \vdots \\ C_n(x,z) - C_n(x,y) \end{bmatrix}$$

- The **consequent difference vectors** are defined analogously, as pitting the consequent mapping $(\widehat{x}, \widehat{y})$ against one of its losers $(\widehat{x}, \widehat{z})$

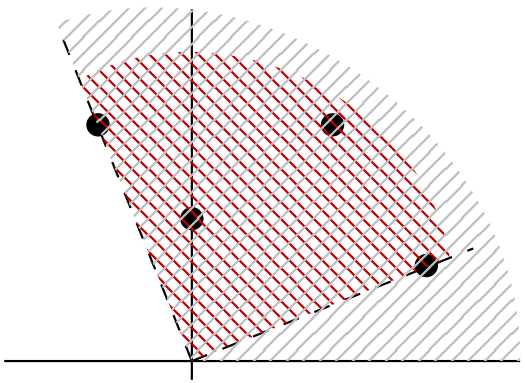
- Suppose there are only $n = 2$ constraints:
  - the antecedent difference vectors can be plotted as points in the plane



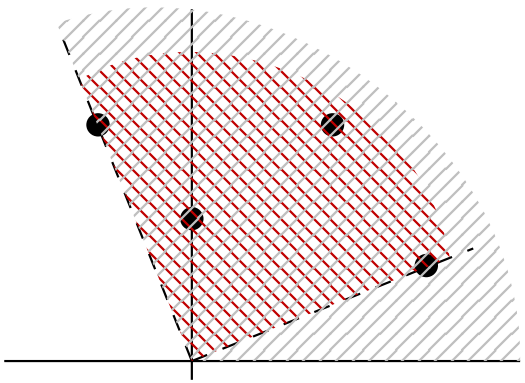
- Suppose there are only $n = 2$ constraints:
  - the antecedent difference vectors can be plotted as points in the plane
  - The red region is their corresponding **cone**



- Suppose there are only $n = 2$ constraints:
  - the antecedent difference vectors can be plotted as points in the plane
  - The red region is their corresponding **cone**
  - The points which are larger than some point in this cone yield the gray region

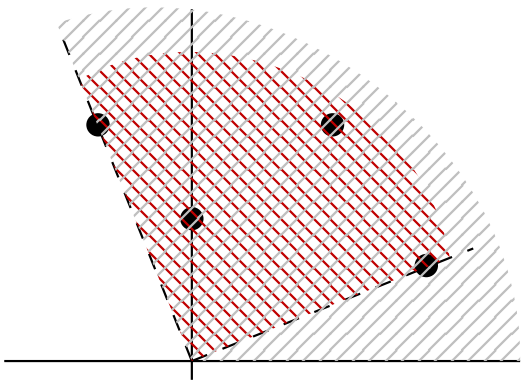


- Suppose there are only $n = 2$ constraints:
  - the antecedent difference vectors can be plotted as points in the plane
  - The red region is their corresponding **cone**
  - The points which are larger than some point in this cone yield the gray region



- The HG T-order $(x, y) \xrightarrow{\text{HG}} (\widehat{x}, \widehat{y})$ holds (for any n) iff each consequent difference vector lives in this gray region: it is larger than some vector in the cone generated by the antecedent difference vectors

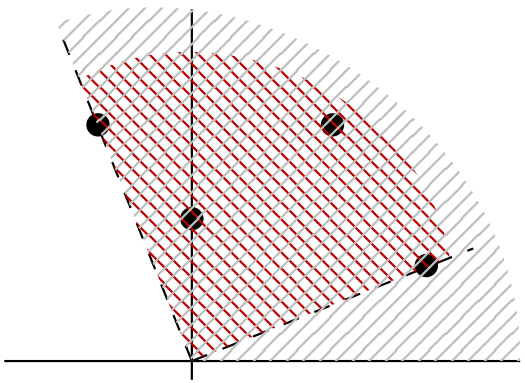
- Suppose there are only $n = 2$ constraints:
  - the antecedent difference vectors can be plotted as points in the plane
  - The red region is their corresponding **cone**
  - The points which are larger than some point in this cone yield the gray region



- The HG T-order $(x, y) \xrightarrow{\text{HG}} (\widehat{x}, \widehat{y})$ holds (for any n) iff each consequent difference vector lives in this gray region: it is larger than some vector in the cone generated by the antecedent difference vectors

- Follows from the *Hyperplane Separation theorem* of convex geometry through straightforward algebra [Boyd and Vandenberghe 2004]

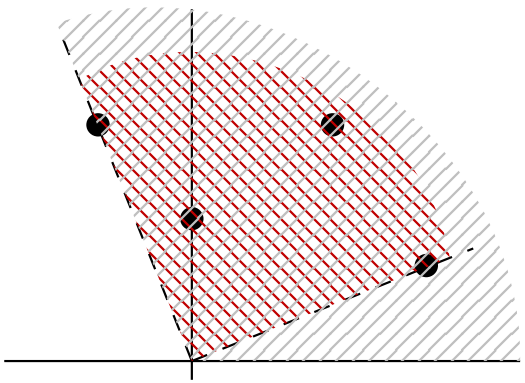
- Suppose there are only $n = 2$ constraints:
  - the antecedent difference vectors can be plotted as points in the plane
  - The red region is their corresponding **cone**
  - The points which are larger than some point in this cone yield the gray region



- The HG T-order $(x, y) \xrightarrow{\text{HG}} (\widehat{x}, \widehat{y})$ holds (for any n) iff each consequent difference vector lives in this gray region: it is larger than some vector in the cone generated by the antecedent difference vectors

- Follows from the *Hyperplane Separation theorem* of convex geometry through straightforward algebra [Boyd and Vandenberghe 2004]
- While $(x, y) \xrightarrow{\text{HG}} (\widehat{x}, \widehat{y})$ is **expensive** to check directly (universal quantification over weights), the geometric characterization above is **easy** to check (polyhedral feasibility problem)

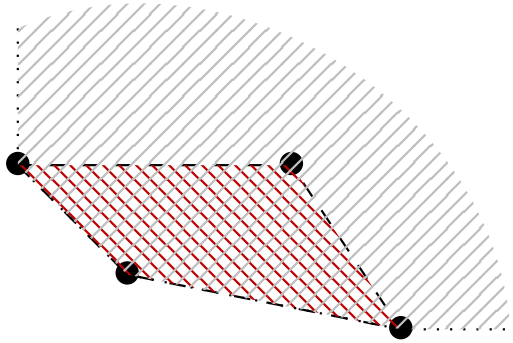
- Suppose there are only $n = 2$ constraints:
  - the antecedent difference vectors can be plotted as points in the plane
  - The red region is their corresponding **cone**
  - The points which are larger than some point in this cone yield the gray region



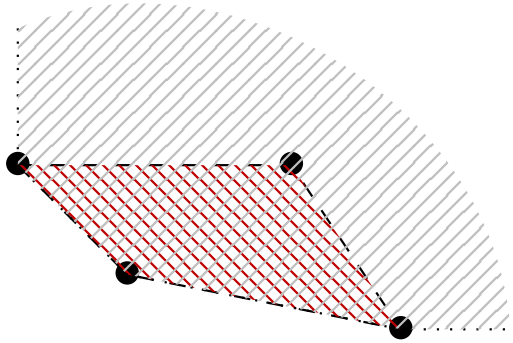
- The HG T-order $(x, y) \xrightarrow{\text{HG}} (\widehat{x}, \widehat{y})$ holds (for any n) iff each consequent difference vector lives in this gray region: it is larger than some vector in the cone generated by the antecedent difference vectors

- Follows from the *Hyperplane Separation theorem* of convex geometry through straightforward algebra [Boyd and Vandenberghe 2004]
- While $(x, y) \xrightarrow{\text{HG}} (\widehat{x}, \widehat{y})$ is **expensive** to check directly (universal quantification over weights), the geometric characterization above is **easy** to check (polyhedral feasibility problem)
- We thus advertise new `Python` code to establish HG T-orders

- Suppose again there are only $n = 2$ constraints:
  - the antecedent difference vectors can be plotted as points in the plane
  - The red region is their **polyhedron**
  - The points which are larger than some point in this polyhedron yield the gray region



- Suppose again there are only $n = 2$ constraints:
  - the antecedent difference vectors can be plotted as points in the plane
  - The red region is their **polyhedron**
  - The points which are larger than some point in this polyhedron yield the gray region

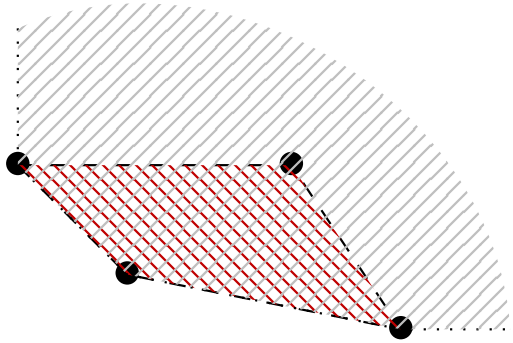


- If the ME T-order $(x, y) \xrightarrow{\text{ME}} (\widehat{x}, \widehat{y})$ holds (for any n), then:

  C1: each consequent difference vector lives in this gray region: it is larger than some vector in the polyhedron generated by the antecedent difference vectors

  C2: the number of candidates of the antecedent ur x is at least as large as the number of candidates of the consequent ur $\widehat{x}$

- Suppose again there are only $n = 2$ constraints:
  - the antecedent difference vectors can be plotted as points in the plane
  - The red region is their **polyhedron**
  - The points which are larger than some point in this polyhedron yield the gray region



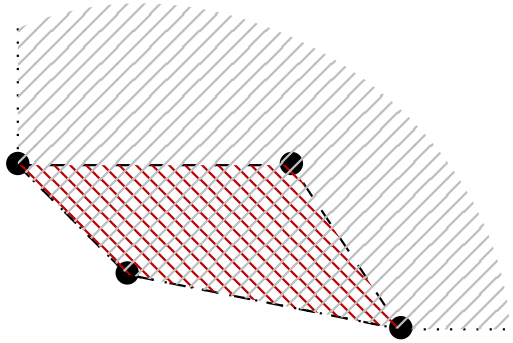
- If the ME T-order $(x, y) \xrightarrow{\text{ME}} (\widehat{x}, \widehat{y})$ holds (for any n), then:

  C1: each consequent difference vector lives in this gray region: it is larger than some vector in the polyhedron generated by the antecedent difference vectors

  C2: the number of candidates of the antecedent ur x is at least as large as the number of candidates of the consequent ur $\widehat{x}$

- C1 is also sufficient when both x and $\widehat{x}$ have at most three candidates

- Suppose again there are only $n = 2$ constraints:
  - the antecedent difference vectors can be plotted as points in the plane
  - The red region is their **polyhedron**
  - The points which are larger than some point in this polyhedron yield the gray region



- If the ME T-order $(x, y) \xrightarrow{\text{ME}} (\widehat{x}, \widehat{y})$ holds (for any n), then:

  C1: each consequent difference vector lives in this gray region: it is larger than some vector in the polyhedron generated by the antecedent difference vectors

  C2: the number of candidates of the antecedent ur x is at least as large as the number of candidates of the consequent ur $\widehat{x}$

- C1 is also sufficient when both x and $\widehat{x}$ have at most three candidates
- We derive a more involved sufficient condition which is nonetheless stronger than C1+C2

Phonological applications 1

Geminate devoicing in Japanese

[Kawahara 2006; Pater 2016]

Lyman's Law: One voiced obstruent per word

Recent loanwords violate Lyman's Law:

(a) [bobu] 'Bob' [giga] 'giga'

Voiced geminates are allowed in recent loanwords:

(b) [web:u] 'web' cf. [wip:u] 'whipped (cream)'

Singleton + geminate voiced obstruent yield optional devoicing:

(c) [gud:o] ~ [gut:o] 'good'
[dog:u] ~ [dok:u] 'dog'

Constraints for Japanese (Pater 2016)

IDENT-VOICE	Assign a violation for an output segment that differs from its input correspondent in [voice].
OCP-VOICE	Assign a violation for a two voiced obstruents within the same word.
*VCE-GEM	Assign a violation for a voiced obstruent geminate.

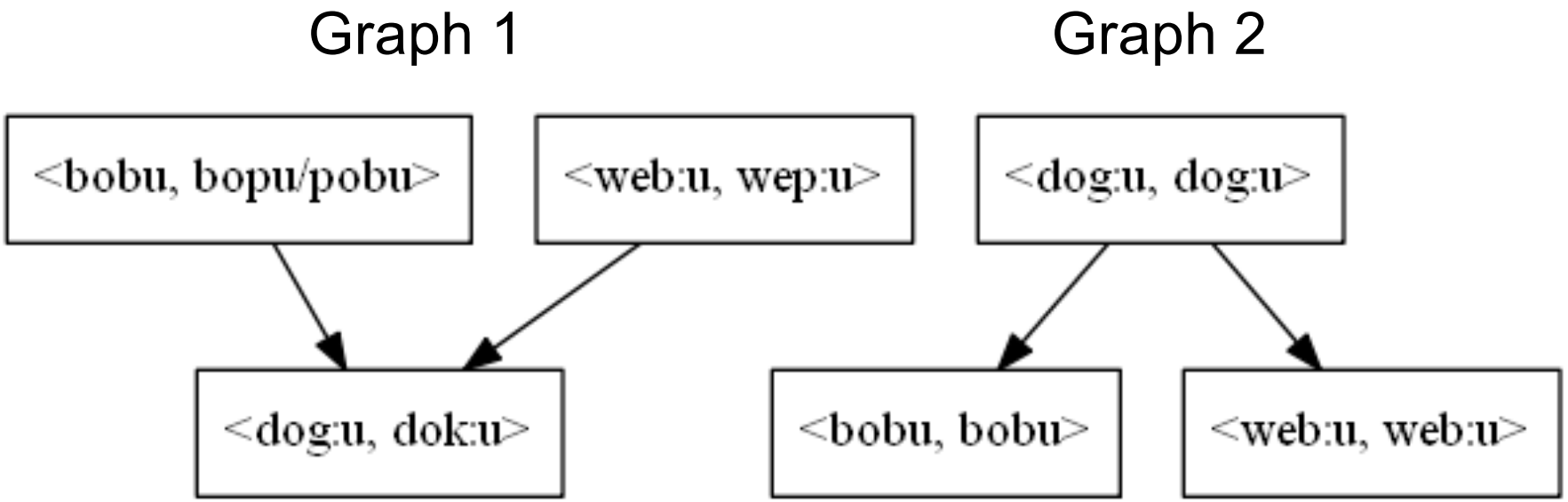
Geminate devoicing in HG (Pater 2016)

	1.5	1	1	Harmony
/bobu/	IDENT-VOICE	OCP-VOICE	*VCE-GEM	
(1a) ☞ bobu		−1		−1
(1b) bopu/pobu	−1			−1.5
/web:u/	IDENT-VOICE	OCP-VOICE	*VCE-GEM	
(2a) ☞ web:u			−1	−1
(2b) wep:u	−1			−1.5
/dog:u/	IDENT-VOICE	OCP-VOICE	*VCE-GEM	
(3a) dog:u		−1	−1	−2
(3b) ☞ dok:u	−1			−1.5

Predicted typology. OT languages are framed.

OT/HG	OT/HG	OT/HG	OT/HG	HG
L1	L2	L3	L4	L5
bobu	bopu/pobu	bobu	bopu/pobu	bobu
web:u	web:u	wep:u	wep:u	web:u
dog:u	dok:u	dok:u	dok:u	dok:u

T-order: OT = HG


Graph 1
Graph 2
<bobu, bopu/pobu>
<web:u, wep:u>
<dog:u, dog:u>
<dog:u, dok:u>
<bobu, bobu>
<web:u, web:u>

Theoretical result

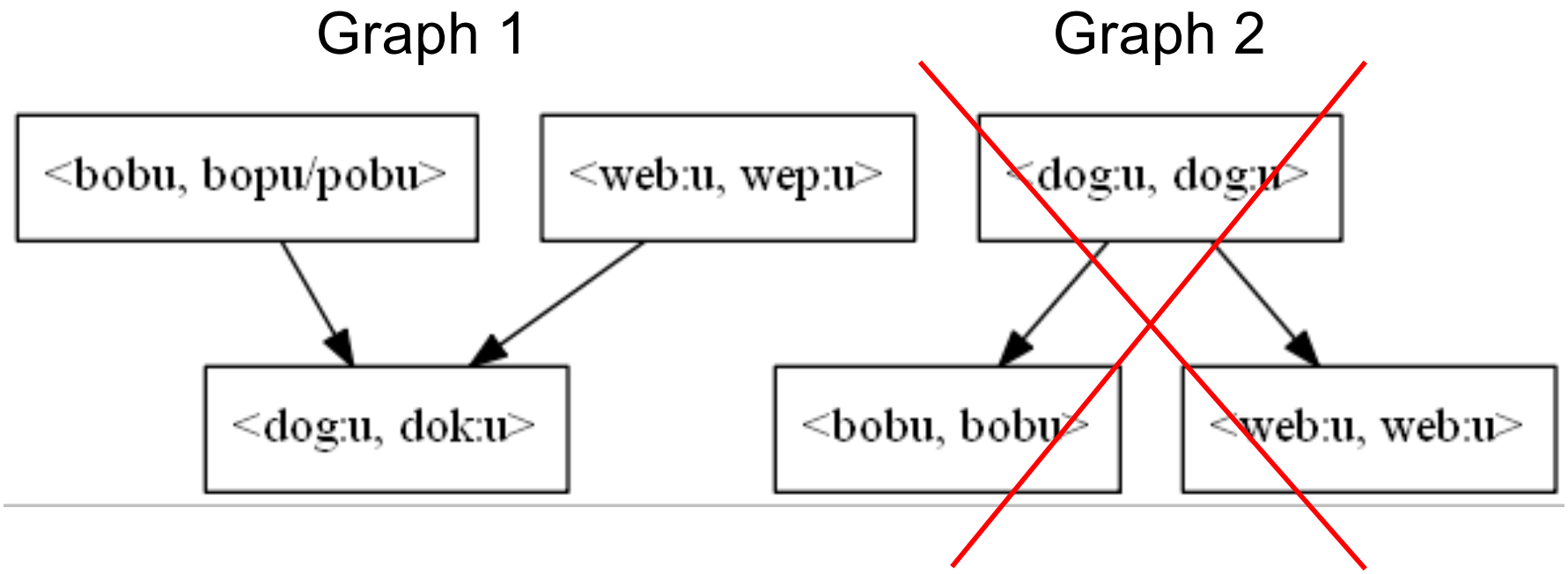
OT and HG T-orders can be identical even when the typologies are different.

- In particular, OT and HG T-orders are identical if there are **only two candidates** (Japanese).
- But this does not exhaust all such cases (e.g., Swedish obstruent voicing, Lombardi 1999, work in progress)

Let's add two harmonically bounded candidates

	1.5	1	1	Harmony
/bobu/	IDENT-VOICE	OCP-VOICE	*VCE-GEM	
(1a) ☞ bobu		−1		−1
(1b) bopu/pobu	−1			−1.5
/web:u/	IDENT-VOICE	OCP-VOICE	*VCE-GEM	
(2a) ☞ web:u			−1	−1
(2b) wep:u	−1			−1.5
/dog:u/	IDENT-VOICE	OCP-VOICE	*VCE-GEM	
(3a) dog:u		−1	−1	−2
(3b) ☞ dok:u	−1			−1.5
(3c) tog:u	−1		−1	−2.5
(3d) tok:u	−2			−3

Maxent T-order



Theoretical result

- In Maxent T-orders, an input with fewer candidates cannot entail one with more candidates.

  Even harmonically bounded candidates count: adding or subtracting them affects the T-order.

Phonological applications 2

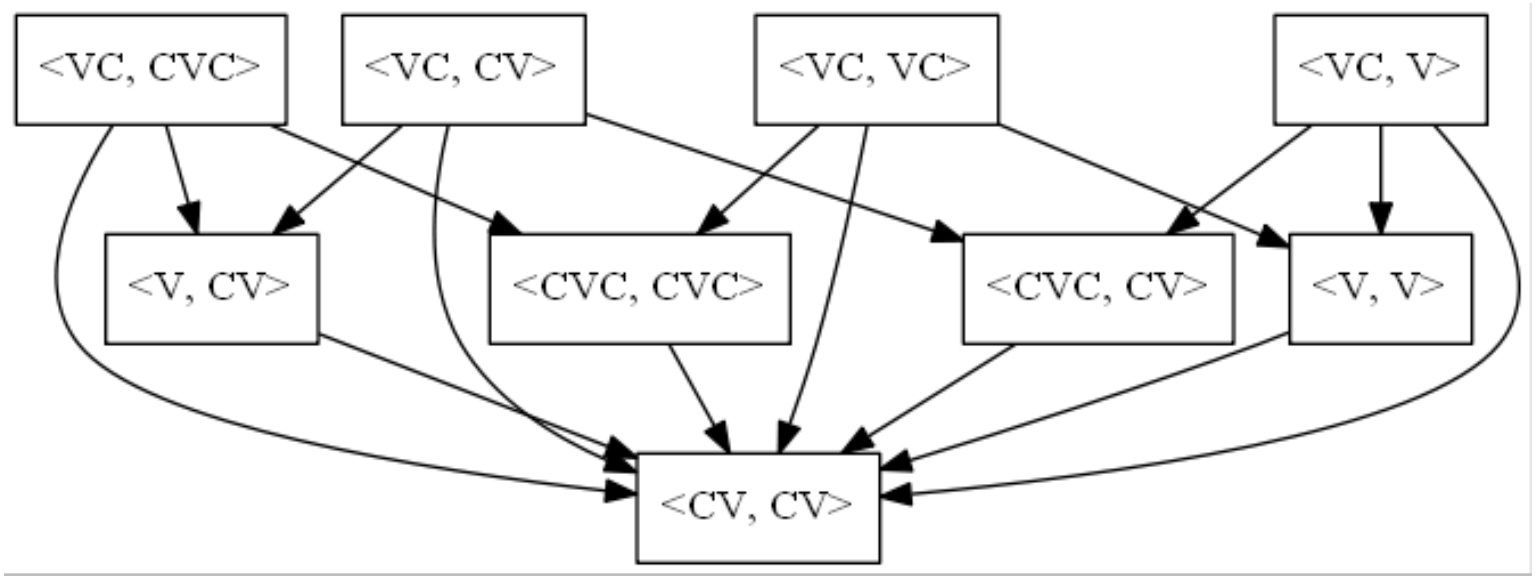
CV syllabification

[Prince and Smolensky 2004]

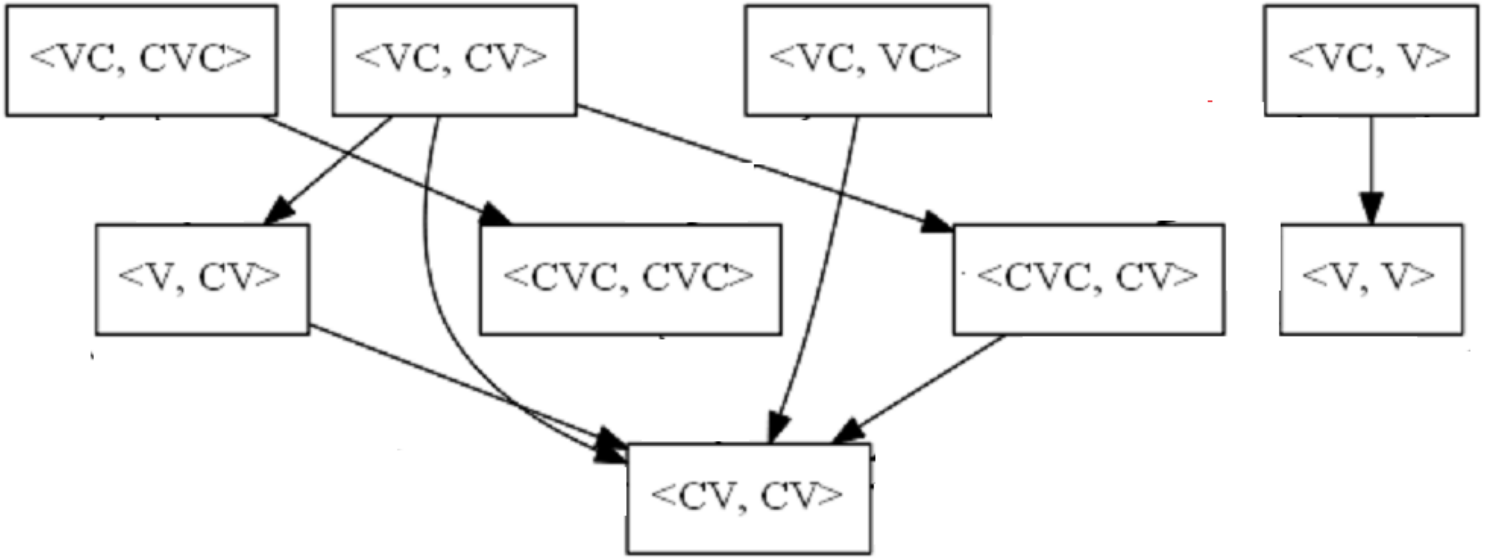
CV syllabification

Input	Output	ONSET	*CODA	MAX	DEP
1. CV	CV				
	CVC		1		1
	VC	1	1	1	1
	V	1		1	
2. CVC	CV			1	
	CVC		1		
	VC	1	1	1	
	V	1		2	
3. VC	CV			1	1
	CVC		1		1
	VC	1	1		
	V	1		1	
4. V	CV				1
	CVC		1		2
	VC	1	1		1
	V	1			

OT/HG T-order



Maxent T-order



What does this mean?

Example: The fact that (CVC, CVC) → (CV, CV) does not hold in Maxent means that there are Maxent grammars where the faithful mapping (CV, CV) has **a smaller probability** than the faithful mapping (CVC, CVC)!

The same holds of the following entailments, which are true in OT/HG, but not in Maxent:

(VC, V) → (CV, CV)
(VC, CVC) → (CV, CV)
(V, V) → (CV, CV)

We now know the following

- OT, HG, and Maxent all have T-orders.
- OT and HG T-orders are identical in certain special cases.
- T-orders can be efficiently computed for all three theories, for HG and Maxent for the first time.
- Maxent T-orders behave in a counterintuitive way: they are sensitive to the **number of candidates**, including harmonically bounded ones.
- Given the standard CV syllabification constraints, Maxent predicts a non-Jakobsonian syllable typology.

Thank you!

Summary: OT vs. HG

Typology	T-orders	Cands	Example
same	same	2	Finnish vowel harmony
same	same	>2	CV syllable typology
same	different	2	CANNOT EXIST
same	different	>2	CANNOT EXIST
different	same	2	Japanese geminate devoicing
different	same	>2	Swedish obstruent devoicing
different	different	2	CANNOT EXIST
different	different	>2	Spanish obstruent lenition/fortition

References

Anttila, Arto. 2008. Gradient phonotactics and the complexity hypothesis. *Natural Language and Linguistic Theory* 26:695–729.

Anttila, Arto, and Curtis Andrus. 2006. T-orders. manuscript and software.

Becker, Michael, Nihan Ketrez, and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in turkish laryngeal alternations. *Language* 87:1:84–125.

Boyd, Stephen, and Lieven Vandenberghe. 2004. *Convex optimization*. Cambridge University Press.

Coetzee, Andries W. 2004. What it means to be a loser: Non-optimal candidates in optimality theory. Doctoral Dissertation, University of Massachusetts, Amherst.

Greenberg, Joseph H. 1963. *Universals of language*. Cambridge, MA: MIT Press.

Guy, G. 1991. Explanation in variable phonology. *Language Variation and Change* 3:1–22.

Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82:536–574.

Kiparsky, Paul. 1993. An ot perspective on phonological variation. Handout, available at http://www. stanford. edu/ kiparsky/Papers/nwave94.

Pater, Joe. 2016. Universal grammar with weighted constraints. In *Harmonic grammar and harmonic serialism*, ed. Joe Pater and John J. McCarthy, 1–46. London: Equinox.

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.