Getting weaker across layers: The tonal phonology of Shona without stratal re-ranking

Eva Zimmermann, joint work with Jochen Trommer

University of Warsaw, linguistics colloquium
June 17th, 2022

Main claim

- Harmonic Layer Theory where phonological elements can get incrementally stronger/weaker at every optimization cycle predicts inter-stratal conspiracies from **a single phonological grammar**

- The theory is **more restrictive** than alternatives based on multiple grammars within a language and makes testable empirical predictions:
  - P1 Monotonicity of phonological changes across strata
  - P2 Consistency of strength in a given stratum
  - P3 Pervasiveness (and cyclicity) of Cooperation

- It further strengthens the arguments for **Gradient Symbolic Representations** in phonology.

Plan

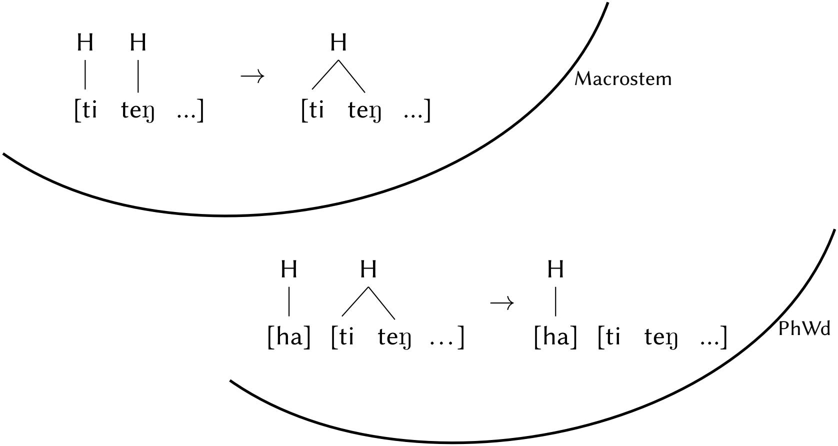
1. Shona: A challenge for a single phonology?

2. Harmonic Layer Theory
2.1 Background assumptions
2.2 A HLT account of Shona

3. Discussion

Shona: A challenge for a single phonology?

The riddle in a nutshell: Inter-stratal conspiracies in Shona

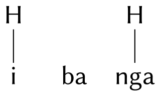


➙ the same marked structure – adjacent H's – is **resolved differently in different morphological contexts**

Background on Shona

- a Bantu language spoken in Zimbabwe
- all data taken from the Zezuni dialect and taken from **?** and **?**
- syllables (=the tone-bearing unit; TBU) can be high-toned (=V́) or low-toned (=V)
- L-tones are taken to be (underlyingly) absent/inserted later

(1) í bangá
'(it) is a hoe'



H H

i ba nga

Domains in Shona

- different **morpho-syntactic domains** are relevant for the phonology

(2) Domains in verbal units, given in ?

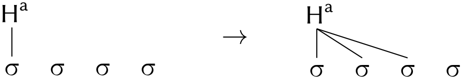
(stem)	root+suffixes
① [macrostem]	optional prefixes (Obj, $\text{Subj/Tns}_{\text{Subj/Part/Neg}}$)+stem
② {phonological word}	optional clitics (e.g. copula, remaining inflection)+macrostem
③ phrase	

(3) {[há]-[ti-(teng-es-e)]}
HORT-1PL/SUBJ-buy-CAUS-FV
'let us sell' (?, 870)

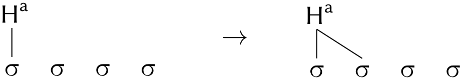
(4) {[ku]-[(téng-és-á)]} {[(sádza)]}
INF-buy-CAUS-FV porridge
'to sell porridge' (?, 862)

The relevant phonological processes: Avoidance of tone-less (=L-toned) TBU's

(5) Spreading to two following TBUs (=H2S)

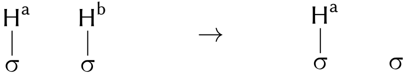


(6) Spreading to one following TBU (=H1S)

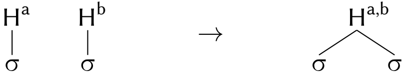


The relevant phonological processes: Avoidance of two adjacent H-tones (=OCP)

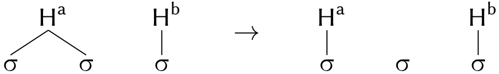
(7) Deletion of the second H (=Del)



(8) Fusion into one (=Fus)



(9) Retraction of a multiply associated first tone (=Retr)



Stratal Differences: Overview

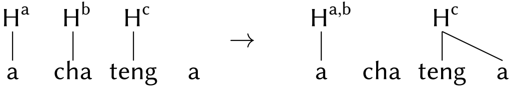
(10)

	H-spread	OCP avoided by:	
① [Macrostem]	H2S	No H-spreading/Retr >	Fus
② {PhWd}	H1S	No H-spreading/Retr >	Del
③ Phrase	H1S	No H-spreading/Retr >	tolerated

Avant: Notation

- underlying H-tones are notated with v̲, surface H-tones with v́

(11) á̲-cha̲-té̲ng-á



Illustrating examples: Spreading

(12) H2S at ①, triggered by Obj
{[tí̲-tárís-e]}
1PL/SUBJ-look-FV
‘we would look’ (**?**, 870)

(13) H1S at ②; triggered by clitic copula
{[í̲]-[sádza]}
COP-porridge
‘(it) is porridge’ (**?**, 860)

(14) H2S at ① and subsequent H1S at ③
{[ku]-[té̲ng-és-á]} {[sádza]}
INF-buy-CAUS-FV porridge
‘to sell porridge’ (**?**, 862)

Illustrating examples: Avoidance of OCP by non-spreading

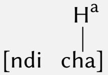
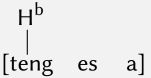
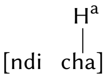
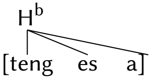
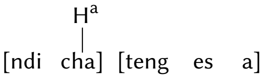
(15) H1S at ②; triggered by clitic copula
{[í̲]-[sádza]}
COP-porridge
'(it) is porridge' (**?**, 860)

(16) H1S at ② blocked if OCP would result
{[í̲]-[badzá̲]}
COP-hoe
'(it) is a hoe' (**?**, 860)

Illustrating examples: Avoidance of OCP by Del

(17) Del at ②
{[ndi-chá̲]-[te̲ng-es-a]}
1.SG-FUT-buy-CAUS-FV
‘I will sell’ (?, 856)

(18)

		underlying representations
		①: Two macrostems
		②: One PhWd

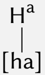
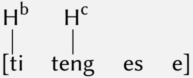
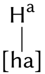
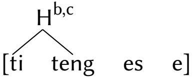
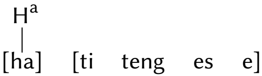
Illustrating examples: Avoidance of OCP by Fus

(19) Fus at ①
{[ku]-[mú̲-té̲ng-és-ér-a]}
INF-OBJ-buy-CAUS-applied-FV
‘to sell him/her’ (**?**, 869)

Illustrating examples: Avoidance of OCP by Fus+Del

(20) Del at ②, fed by Fus at ①
{[há]-[ti-tenges-e]}
HORT-1PL/SUBJ-buy-CAUS-FV
‘let us sell’ (?, 870)

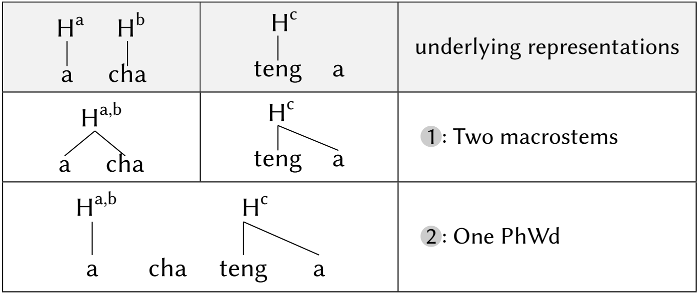
(21)

		underlying representations
		①: Two macrostems
		②: One PhWd

Illustrating examples: Avoidance of OCP by Fus+Retr

(22) Retr at ②, fed by Fus at ①
{[á-cha]-[téng-á]}
3SG-FUT=buy-FV
‘s/he will buy’ (?, 864)

(23)



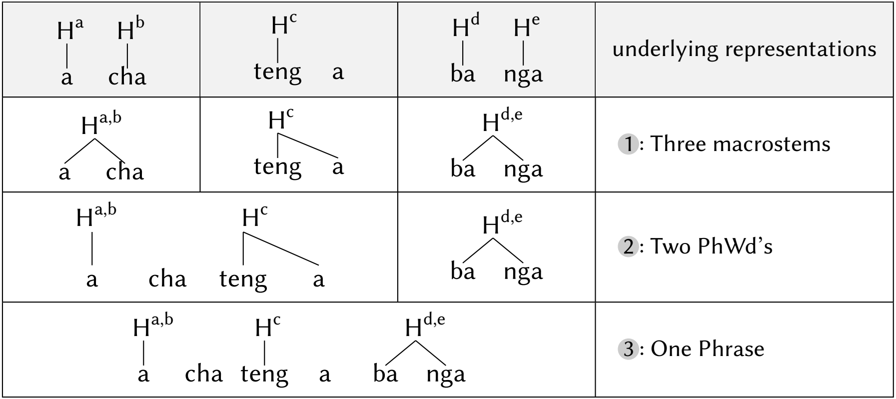
underlying representations

①: Two macrostems

②: One PhWd

Interaction of processes at different layers: More complex example

(24)



underlying representations

1: Three macrostems

2: Two PhWd's

3: One Phrase

Illustrating examples: OCP cannot be avoided

(25) OCP tolerated if Retr impossible at ③
{[badzá̲]} {[gú̲rú]}
hoe big
‘big hoe’ (**?**, 874, FN.21)

Summary: Stratal Differences

(26)

	H-spread	OCP avoided by:	
① [Macrostem]	H2S	No H-spreading/Retr >	Fus
② {PhWd}	H1S	No H-spreading/Retr >	Del
③ Phrase	H1S	No H-spreading/Retr >	tolerated

Harmonic Layer Theory

Background assumptions

Harmonic Layer Theory: Overview

- phonological evaluations at every **morphological layer**
- linguistics elements have **gradient activity** that results in gradient constraint violations (Gradient Symbolic Representations; =GSR)
- tones can **get stronger or weaker in every layer** and the ‘same’ tone can react differently to identical tonotactic problems in larger domains since it has different activity

→ different phonological behaviour results from a **single phonological grammar**

(=vs. stratal model (**????**) with optimizations at every stratum with a potentially *different grammar*)

Background: Gradient Symbolic Representations (=GSR)

- all linguistic symbols have **activity** that can **gradiently** differ and 1 is the default activity (**??**)
- any change in activity is a faithfulness violation – different activities result in **gradient violations of faithfulness**
- elements can be gradiently active in the output and thus violate **markedness constraints gradiently**
(**??????**)
- grammatical computation modeled inside **Harmonic Grammar** where constraints are weighted (**??**)

GSR and constraint violations

- constraints are **violated/satisfied relative to the activity** of the relevant elements
- elements preferably have the default activity of 1 (=*Weak, *Strong)
- e.g. the underlyingly weakly active segment in (27)
  - is **easier to delete** than a fully active segment
  - is **costly to realize**
  - **tolerates more marked structures**

(27) Gradient activity=gradient constraint violations

$b_1a_1t_1$-$p_{0.5}$	*Weak 10	MaxS 10	DepS 10	*CC 10		
a. $b_1a_1t_1p_1$			**-0.5**	**-1**	-15	Only fully active S
b. $b_1a_1t_1p_{0.5}$	**-0.5**			**-0.75**	-12.5	Faithful realization of weak S
c. $b_1a_1p_{0.5}$	**-0.5**	**-1**			-15	Deletion of fully active S
☞ d. $b_1a_1t_1$		**-0.5**			-5	Deletion of weakly active S

GSR: Broader Context

- that linguistic elements are not categorical but have strength differences is **not a new** idea

  (e.g. **?** and **?** for functional categories in syntax, **?**: some lexical accent system are based on scalar grades of accent strength,...)
- other work on non-categorical elements in **neural networks**

  (e.g. **?** on induction of prosodic categories in neural networks)
- can also predict **phonetic gradience**

  (e.g. subphonemic gradience in word-final devoicing, nasal place assimilation, flapping (e.g. **?**), vowel harmony is gradient (**?**),...)
- **different from a binary** distinction into strong/weak
  (**????**)

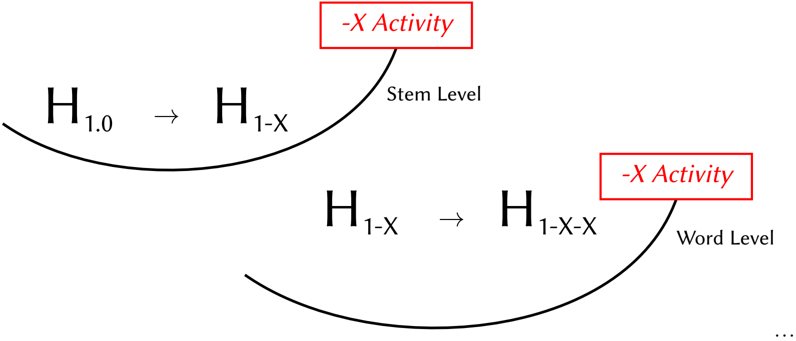
→ here: predictions of **gradient (=numerical) phonological strength in an OT-system** as explanation for ‘exceptional’ behaviour

General Arguments for GSR

1. Embedded in a general **computational architecture for cognition**
   (=Gradient Symbolic Computation, **?**)
2. A **unified account** for different exceptional phonological behaviours:
   - liaison consonants in French (**?**)
   - semi-regularity of voicing in Japanese Rendaku (**?**)
   - allomorphy in Modern Hebrew (**?**)
   - lexical accent in Lithuanian (**?**)
   - tone sandhi in Oku (**?**)
   - tone allomorphy in San Miguel el Grande Mixtec (**??**)
   - lexical stress in Moses Columbian Salishan (**?**)
   - exceptional tone (non)spreading in San Molinos Mixtec (**?**)
   - interaction of phonological/lexical gemination/lenition in Italian (**?**)
   - compound stress in Sino-Japanese (**?**)
   - (interacting) ghost segments in Welsh (**?**)
   - …

HLT: Predictable loss/gain of activity at every layer

- constraint interaction can ensure that all instances of a certain element (e.g. H) **gain or loose a fixed amount of activity** at every optimization cycle



-X Activity

$\mathrm{H}_{1.0} \rightarrow \mathrm{H}_{1\text{-}X}$

Stem Level

-X Activity

$\mathrm{H}_{1\text{-}X} \rightarrow \mathrm{H}_{1\text{-}X\text{-}X}$

Word Level

…

A HLT account of Shona

Activity loss at every stratum

- constraint interaction ensures that all **H’s decay** by 0.2 at every layer

(28) Predictable decay by 0.2

	INPUT	OUTPUT
①	H_1	$H_{0.8}$
②	$H_{0.8}$	$H_{0.6}$
③	$H_{0.6}$	$H_{0.4}$

(29) $^*\Sigma_H$: Assign -x violation for every H_x.

(30) $|\Delta\mathcal{S}| \leq 0$: Assign -x violation for every input tone H_a corresponding to output tone H_b where a-b=x and x is > 0.

(31) $|\Delta\mathcal{S}| \leq 0.2$: Assign -x violation for every input tone H_a corresponding to output tone H_b where a-b=x and x is > 0.2.

Shona HLT account: Decrease of H-tone activation

(32) Macrostem level: $H_{1.0} \rightarrow H_{0.8}$

$H_{1.0}$	$\lvert\Delta\mathcal{S}\rvert \leq 0.2$ w=∞	Max H w=11	$^*\Sigma_{\text{H}}$ w=10	$\lvert\Delta\mathcal{S}\rvert \leq 0$ w=1	$\mathcal{H}$
a. $H_{1.0}$			-1.0		-10
b. Ø		-1.0			-11
c. $H_{0.5}$	-0.5		-0.5	-0.5	∞
☞ d. $H_{0.8}$			-0.8	-0.2	-8.2

(33) PhWd level: $H_{0.8} \rightarrow H_{0.6}$

$H_{0.8}$	$\lvert\Delta\mathcal{S}\rvert \leq 0.2$ w=∞	Max H w=11	$^*\Sigma_{\text{H}}$ w=10	$\lvert\Delta\mathcal{S}\rvert \leq 0$ w=1	$\mathcal{H}$
a. $H_{0.8}$			-0.8		-8
b. Ø		-0.8			-8.8
c. $H_{0.5}$	-0.3		-0.5	-0.3	∞
☞ d. $H_{0.6}$			-0.6	-0.2	6.2

Different behaviour for spreading: In a nutshell

- providing a TBU with a tone to avoid a violation of Spec gets less helpful, the weaker the tone is

(34)

INPUT **OUTPUT**

① H2S H_1 - V́ V V → $H_{0.8}$ – V́ V́ V́
➞ 0.8xSpec > *H_{3TBU}

② H2S $H_{0.8}$ - V́ V V → $H_{0.6}$ – V́ V́ V́
➞ *H_{3TBU} > 0.6xSpec

③ H1S $H_{0.6}$ - V́ V V → $H_{0.4}$ – V́ V́ V
➞ *H_{3TBU} > 0.4xSpec

Shona HLT account: Constraints I

(35) SPECIFY
Assign -(1-X) violation for every TBU associated with tone T with activity X (and no tone is X=0).

(36) $^{*}H_{2TBU}$
Assign -1 violation for every tone that is associated to more than one TBU.

(37) $^{*}H_{3TBU}$
Assign -1 violation for every tone that is associated to more than two TBU's.

Tableaux: H2S at ① but H1S at ②

(38) ①: H2S

$H_{1.0}$ V V	SPEC 90	$^{*}H_{3TBU}$ 56	$^{*}H_{2TBU}$ 1	$\mathcal{H}$
a. $H_{0.8}$ – V́ V V	-2.2			-198
b. $H_{0.8}$ – V́ V́ V	-1.4		-1.0	-127
☞ c. $H_{0.8}$ – V́ V́ V́	-0.6	-1.0	-1.0	-111

(39) ②: H1S

$H_{0.8}$ V V	SPEC 90	$^{*}H_{3TBU}$ 56	$^{*}H_{2TBU}$ 1	$\mathcal{H}$
a. $H_{0.6}$ – V́ V V	-2.4			-216
☞ b. $H_{0.6}$ – V́ V́ V	-1.8		-1.0	-163
c. $H_{0.6}$ – V́ V́ V́	-1.2	-1.0	-1.0	-165

Tableaux: H1S at ② and ③

(40) ②: H1S, repeated

$H_{0.8}$ V V	Spec 90	*H_{3TBU} 56	*H_{2TBU} 1	$\mathcal{H}$
a. $H_{0.6}$ – V́ V V	-2.4			-216
☞ b. $H_{0.6}$ – V́ V́ V	-1.8		-1.0	-163
c. $H_{0.6}$ – V́ V́ V́	-1.2	-1.0	-1.0	-165

(41) ③: H1S

$H_{0.6}$ V V	Spec 90	*H_{3TBU} 56	*H_{2TBU} 1	$\mathcal{H}$
a. $H_{0.4}$ – V́ V V	-2.6			-234
☞ b. $H_{0.4}$ – V́ V́ V	-2.2		-1.0	-199
c. $H_{0.4}$ – V́ V́ V́	-1.8	-1.0	-1.0	-219

Different behaviour for OCP problems: In a nutshell

- the weaker the H, the cheaper deletion (and the more costly fusion)
- the weaker the H, the easier it is to tolerate the OCP

(42) INPUT OUTPUT

① Fusion $H_1 + H_1 \rightarrow (H_{0.8}H_{0.8})$

➙ MAX > UNIF

➙ 0.8xOCP > MAX / UNIF

② Deletion $H_{0.8} + H_{0.8} \rightarrow H_{0.8}$

➙ UNIF > 0.8xMAX

➙ 0.6xOCP > 0.8xMAX / UNIF

③ – $H_{0.6} + H_{0.6} \rightarrow H_{0.6}H_{0.6}$

➙ UNIF / 0.6xMAX > 0.4xOCP

Shona HLT account: Constraints II

(43) MAXT: Assign -x violation for every H_x in the input without an output correspondent.

(44) OCP: Assign -$\frac{x+y}{2}$ violation for every pair of adjacent tones H_x and H_y that are associated with adjacent TBU's.

(45) UNIF: Assign -1 violation for every pair of input tones corresponding to the same output tone.

Tableaux: OCP resolution I

(46) ①: Tone Fusion

$H_{1.0}$ $H_{1.0}$	OCP 23	MaxT 16	Unif 14	$\mathcal{H}$
a. $H_{0.8}$ $H_{0.8}$	-0.8			-18.4
b. $H_{0.8}$		-1.0		-16
☞ c. ($H_{0.8}$ $H_{0.8}$)			-1.0	-14

(47) ②: Tone Deletion

$H_{0.8}$ $H_{0.8}$	OCP 23	MaxT 16	Unif 14	$\mathcal{H}$
a. $H_{0.6}$ $H_{0.6}$	-0.6			-13.8
☞ b. $H_{0.6}$		-0.8		-12.8
c. ($H_{0.6}$ $H_{0.6}$)			-1.0	-14

Tableaux: OCP resolution II

(48) ②: Tone Deletion, repeated

$H_{0.8}$ $H_{0.8}$	OCP 23	MaxT 16	Unif 14	$\mathcal{H}$
a. $H_{0.6}$ $H_{0.6}$	-0.6			-13.8
☞ b. $H_{0.6}$		-0.8		-12.8
c. ($H_{0.6}$ $H_{0.6}$)			-1.0	-14

(49) ③: OCP violation tolerated

$H_{0.6}$ $H_{0.6}$	OCP 23	MaxT 16	Unif 14	$\mathcal{H}$
☞ a. $H_{0.4}$ $H_{0.4}$	-0.4			-9.2
b. $H_{0.4}$		-0.6		-9.6
c. ($H_{0.4}$ $H_{0.4}$)			-1.0	-14

HLT account of Shona

- loosing 0.2 activity at each optimization predicts the different phonological behaviours in Shona from a **single grammar**

Discussion

Predictions of HLT

- in contrast to accounts based on multiple grammars, HLT makes several testable predictions:

  P1 ***Monotonicity*** of phonological changes across strata

  P2 **Consistency** of strength in a given stratum

  P3 Pervasiveness (and cyclicity) of **Cooperation**

P1: Monotonicity

Representations become monotonically stronger or weaker
+ single constant grammar
= monotonicity of phonological behaviour

(50) Monotonicity of thresholds for phonological behavior in HLT

T_x ➙ Phonological behavior 1

WEAKER: THRESHOLD 1

$T_{x\text{-}y}$ ➙ Phonological behavior 2

WEAKER: THRESHOLD 2

$T_{x\text{-}y\text{-}z}$ ➙ Phonological behavior 3

P1: Monotonicity in Shona

(51) The Shona pattern

	OCP:	H-spread
① [Macrostem]	Fus	H2S
② {PhWd}	Del	H1S
③ Phrase	tolerated	H1S

(52) Impossible in HLT

	OCP:	H-spread
① [Macrostem]	Fus	H2S
② {PhWd}	Del	H1S
③ Phrase	Fus	H1S

(53) No monotonicity with stratum-specific rankings

Macrostem Level: MaxH $\gg$ OCP
PhWd Level: OCP $\gg$ MaxH
Phrase Level: MaxH $\gg$ OCP

P2: Consistency of strength

Different repairs for elements must be contingent with their input strength since constraint weighting remains constant.

(54) Consistency-obeying: Giphende Nominal Morphology

Citation Form:	a.	L-LL	b.	L-LH	c.	L-HL	d.	L-HH
Focus:		H-HL		L-LH		L-HL		L-HH
Genitive:		H-HL		H-LH		L-HL		L-HH
Predicative:		H-HL		H-LH		H-HL		H-HH

(55) Consistency-violating: Construction-specific rankings

		H$]_{\text{PrWd}}$	**H**H
Construction 1	$M_1 \gg F \gg M_2$	Deletion	No deletion
Construction 2	$M_2 \gg F \gg M_1$	No Deletion	Deletion

P3: Pervasiveness of Cooperation

Multilateral conditioning of morphophonological processes: Fused phonological material of different strength may contribute cumulatively to phonological behavior

- ➙ Lexical conditioning is the existence of weak elements that need to undergo fusion with another weak element

(56) Cooperation as lexical idiosyncrasy

a. $$\begin{array}{cccccccc} & & H_{0.5} & & & & & \\ \sigma & \sigma & & - & \sigma & \rightarrow & \sigma\ \sigma\ \sigma \end{array}$$

b. $$\begin{array}{cccccccccc} & & H_{0.5} & & H_{0.5} & & & & & H_1 \\ & & & & & & & & & | \\ \sigma & \sigma & & - & & \sigma & \rightarrow & \sigma & \sigma & \sigma \end{array}$$

Summary

- HLT predicts inter-stratal conspiracies as in Shona from a **single grammar** if elements can consistently loose/gain activity at every optimization step
- In contrast to accounts based on multiple grammars, it makes **testable predictions** about possible different behaviours within a language

References I