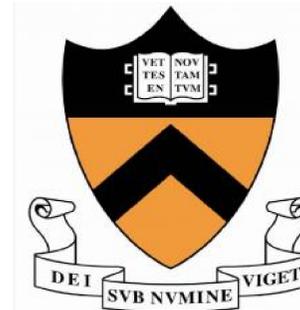# A Diachronic Counter-example to the Subset Principle: The Case of Anatolian Reduplication

Anthony D. Yates, UCLA (adyates@ucla.edu) & Sam Zukoff, Princeton (szukoff@princeton.edu)

1. Introduction

- The **"Subset Principle"** (cf. Prince & Tesar 2004): learners choose the most restrictive grammar consistent with the positive evidence (≈ initial bias of MARKEDNESS ≫ IO-FAITHFULNESS).
  - Capturing the Subset Principle is key argument in favor of Biased Constraint Demotion (BCD; Prince & Tesar 2004) and Low Faithfulness Constraint Demotion (LFCD; Hayes 2004) over simple Recursive Constraint Demotion (RCD; Tesar & Smolensky 1998).

- **The diachronic development of the Anatolian reduplicative system** represents a case where speakers learned a **superset** grammar.
  - → Our learning algorithm needs to accommodate (a specific kind of) non-Subset learning.

- **Proposal: "Maximally Informative Recursive Constraint Demotion" (MIRCD)**
  - A version of RCD (or BCD) which is biased towards winner-preferring constraints that can account for the greatest amount of data possible.
  - Non-Subset learning is permitted with MIRCD when there is a superset-subset relationship between the violation profiles of crucial constraints.

2. Anatolian Data

- **Reduplication patterns** of Hittite and its (reconstructed) proximate ancestor Proto-Anatolian [abbrev. PA] (following Yates & Zukoff 2016a,b, in press, Zukoff 2017; cf. Dempsey 2015):

	Base Shape	Proto-Anatolian	>	Hittite
	CVX–	*CV-CVX–	>	CV-CVX–
(1)	TRVX–	*TV-TRVX–	>	TRV-TRVX–
	STVX–	*STV-STVX–	>	iSTV-STVX–
	VRTX–	does not exist yet	>	VR-VRTX–

★ Distinction in copying patterns in PA between *obstruent-sonorant* (TRVX–) bases and *s-obstruent* (STVX–) bases: **C₁-copying vs. cluster-copying.**

- Distribution analyzed with *PCR, a constraint against certain types of consonant repetitions:

(2) **No Poorly-Cued Repetitions (*PCR)** [≈ *$C_\alpha V C_\alpha$ / _$C_{\text{[-sonorant]}}$] (Zukoff 2017)
For each sequence of repeated identical consonants separated by a vowel ($C_\alpha V C_\alpha$), assign a violation mark * if that sequence immediately precedes an obstruent.

3. Anatolian Analysis & Diachrony

- CONTIGUITY-BR (McCarthy & Prince 1995) must rank low in PA to allow TRVX– C₁-copying in *pi-pri–* (4.i), but high in Hittite to generate TRVX– cluster-copying in *pri-pri–* (4.ii).

- *PCR must rank high in PA to generate STVX– cluster-copying in *stu-stu–* (5.i), **but low in Hittite to allow VRTX– *ar-ark–* to emerge (6.ii).**

- ALIGN-ROOT-L is ranked in the middle at both stages.

(3) **ALIGN-ROOT-L**: Assign one violation mark * for each segment intervening between the left edge of the root and the left edge of the word.

4. Anatolian Analysis & Diachrony (cont.)

i. Proto-Anatolian > ii. Hittite

(4)

/RED, pri-/	*PCR	ALIGN	CNTG_BR
a. ☞ pi-pri–		**	*
b. pri-pri–		***!	

/RED, pri-/	CNTG_BR	ALIGN	*PCR
a. pi-pri–	*!	**	
b. ☞ pri-pri–		***	

(5)

/RED, stu-/	*PCR	ALIGN	CNTG_BR
a. su-stu–	*!	**	*
b. ☞ stu-stu–		***	

/RED, stu-/	CNTG_BR	ALIGN	*PCR
a. su-stu–	*!	**	*
b. ☞ istu-stu–		***	

(6)

No pattern exists yet: no vowel-initial roots in PA (Yates & Zukoff in press)

/RED, ark-/	CNTG_BR	ALIGN	*PCR
a. ☞ ar-ark–		**	*
b. ark-ark–		***!	

- Total demotion of *PCR constitutes a **diachronic counter-example to the Subset Principle**.
  - Learners learned a grammar with a low-ranked markedness constraint despite not encountering evidence that it was violable.

- **Question:** What could have led learners to fail to obey the Subset Principle in this case?

- ★ **Proposal:** **Maximally Informative Recursive Constraint Demotion (MIRCD)**
  - RCD with a preliminary step which picks out and installs the *constraints that favor the most winners* first (cf. Becker 2009), i.e., the most "informative" constraints.

5. MIRCD in Pre-Hittite [after change from TV-TRVX– to TRV-TRVX–]

- MIRCD first installs CNTG_BR because it has only **W**'s, and the most **W**'s.
  - → RCD would install *PCR, but MIRCD does not because it does not have the most W's.

(7)

MIRCD (round 1) ⇒ install CNTG_BR	CNTG_BR	*PCR	ALIGN	MAX_BR
i. TRVCV- → TRV-TRVCV- ≻ TV-TRVCV-	W	e	L	W
ii. TRVCV- → TRV-TRVCV- ≻ TRVCV-TRVCV-	e	e	W	L
i. STVCV- → STV-STVCV- ≻ SV-STVCV-	W	W	L	W
ii. STVCV- → STV-STVCV- ≻ STVCV-STVCV-	e	e	W	L

- Among remaining support (white rows), ALIGN is the only winner-preferrer, so it gets installed.
  - → Again unlike RCD, MIRCD does not install PCR despite it preferring no losers.

- All data is now explained, so *PCR (and MAX_BR) are ranked at the bottom of the grammar.
  - → **This is the ranking necessary to allow the later emergence of VR-VRTX–.**

- Non-Subset learning is permitted w/ MIRCD here because of the superset-subset relationship between CNTG_BR and *PCR: *PCR explains a proper subset of the data which CNTG_BR explains.

- Under these specific conditions, MIRCD produces the non-Subset learning necessary to capture the Anatolian facts, without predicting non-Subset learning in the general case.