Paamese-type Arbitrary-sized Stress Windows are Learnable

(if you are very patient)

Ryan Sandell, Ludwig-Maximilians-Universität München

Manchester Phonology Meeting 31

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

1. Research Questions

1. What is the relationship between **typology** and **learnability** in stress systems?
2. Specifically: are STRESS WINDOWS subject to "hard" or "soft" limits on learnability?
   - **Hard Limits:** Constraints / Parameters or similar furnished by UG strictly exclude conceivable patterns from occurring in natural languages.
   - **Soft Limits:** Certain patterns are theoretically learnable, but are disadvantaged relative to other patterns.
3. **Claim:** there is no theoretical upper bound on the size of learnable stress windows.
   But: systems with larger windows are disadvantaged on account of restrictions on evidence and the learning paradigm.

2. Data: Paamese

1. Primary stress in Paamese (Southern Oceanic; Vanuatu; see Crowley 1982) exhibits an ostensible (reduced) four-syllable stress window (final syllable is always unstressed).
   - See discussion in Goldsmith 1990: 215–16, Hayes 1995: 178–79, Lee 1999, and Kager 2012: 1466

   (1) PAAMESE STRESS ALGORITHM: assign primary stress to the antepenult, unless the antepenult is lexically unstressable (V̆), else to the preantepenult, unless the preantepenult is lexically unstressable, else to the penult.

   (2) PAAMESE STRESS PATTERNS

a.	*i.náu.li.i*	'oh, me'	*i.nau.li.í.ri.si*	'oh, me again'
b.	*sú.u.hi*	'it scrapes'	*na.sú.u.hi*	'I scrape'
c.	*mó.lă.ti.ne*	'man'	*mo.lă.tí.ne.se*	'only the man'
d.	*tă.hó.si*	'it is good'	*ná.tă.ho.si*	'I am good'
e.	*tŏ.vŭ.é.li*	'not exist'		

   - The stress pattern reported by Crowley (1982) is generated at an intermediate level of representation. Subsequent processes (syncope, coalesence, etc.) opacify the predictable stress distributions.
2. **Paamese-type stress system:** default stress does not mark the edge of the stress window; stress prefers to go beyond the default position rather than retract towards the word edge.
3. Is the Paamese four-syllable window learnable?
   Are yet larger stress windows with a Paamese-type pattern driven by lexical properties learnable?

3. Existing Analysis: Lee 1999

1. For the Paamese data in (2): generate a single right-aligned binary trochaic foot that never results in stress on a /V̆/ and that prefers to exclude the final syllable from the foot.

   (3) **Constraint Ranking in Lee 1999:**

   FTBIN, TROCHEE, *[ˈə] ≫ NONFIN-FT ≫ ALL-FT-R ≫ PARSE-σ
2. Overgeneration: this yields an unbounded stress pattern, with primary stress on the rightmost stressable syllable that avoids a violation of NONFIN-Ϝ.
   - A word of five or more syllables containing an unstressable antepenult and preantepenult, /σ₁ σ̆ σ̆ σ σ/, will receive primary stress on the stressable fifth-to-last syllable per (3).
   - The absence of an output like [ˈσ σ̆ σ̆ σ σ] in Paamese must by explained by the grammar, given Richness of the Base.

   (4) Overgeneration of Lee 1999: /σ σ̆ σ̆ σ σ/ → [ˈσ σ̆ σ̆ σ σ]

/σ σ̆ σ̆ σ σ/		TROCHEE	*[ˈə]	NONFIN-Ϝ	ALL-Ϝ-R	PARSE-σ
a.	☞ (ˈσ σ̆) σ̆ σ σ				***	***
b.	σ σ̆ σ̆ (ˈσ σ)			*!		***
c.	σ (ˈσ̆ σ̆) σ σ		*!		**	***
d.	σ σ̆ (ˈσ̆ σ) σ		*!		*	***
e.	σ σ̆ (σ̆ ˈσ) σ	*!				***

- /σ̆/ interpreted as vowel marked as [–stress] (de Lacy 2020)
- *[ˈə] employed by Leee not really a markedness constraint: better understood as IDENT-[stress].

4. Window Restriction and Ganging

1. How to avoid unbounded stress in this context?
   - More *LAPSE or ALL-Ϝ-L/R×N constraints?
   - Constraint (self-)ganging of ALL-Ϝ-L/R
2. ALL-FT-L/R can act as a size restrictor to enforce a stress window (Legendre et al. 2006).
   - In the Paamese case: $3 \times W$(ALL-Ϝ-R) $> W$(NONFIN-Ϝ).
   - Windows of arbitrary size $N + 2$ possible (N = # violations of ALL-Ϝ-R/L).
3. **But:** four-syllable windows rare/disputed, and yet larger windows unattested (Hayes 1995, Goedemans et al. 2014). **Why?**

5. Simulation Preparation

- Data for simulations generated via scripting in R v. 4.4.2.
- Strings consist of stressable (σ) and unstressable (σ̆) syllables, 2–8 syllables in length.

# Inputs	# Overt Forms	# Parses
508	3584	9736
/σ σ̆ σ σ/	[ˈσ σ̆ σ σ]	[(ˈσ σ̆) σ σ]

- Learners are assumed to know which syllables are lexically unstressable; these lexical properties are not learned simultaneously.
- 12 constraints employed (Kager 1999: Ch. 4): ALIGN-L/R(ω, Ϝ), ALL-Ϝ-L/R, FTBIN, TROCHEE, IAMB, PARSE-σ, NONFINALITY-Ϝ/σ, NONINITIALITY-Ϝ, **IDENT-[stress]**.

6. Patterns & 7. Model Parameters

- Expected probabilities of stress positions in the 3584 overt forms for five categorical Paamese-like window systems (2–6 syllables) with default penult or antepenult stress generated.
  2- and 3-syllable windows permit final stress; larger windows do not.

(5) Expected overt outputs for six-syllable inputs

INPUT	WINDOW SIZE 3	4 = Paamese	5	6
/σ σ σ σ σ σ/	[σ σ σ ˈσ σ σ]	[σ σ σ ˈσ σ σ]	[σ σ σ ˈσ σ σ]	[σ σ σ ˈσ σ σ]
/σ σ σ σ σ̆ σ/	[σ σ σ ˈσ σ̆ σ]	[σ σ σ ˈσ σ σ]	[σ σ σ ˈσ σ σ]	[σ σ σ ˈσ σ σ]
/σ σ σ σ̆ σ̆ σ/	[σ σ σ σ̆ σ̆ ˈσ]	[σ σ ˈσ σ̆ σ̆ σ]	[σ σ ˈσ σ̆ σ̆ σ]	[σ σ ˈσ σ̆ σ̆ σ]
/σ σ σ σ̆ σ σ/	[σ σ σ σ̆ ˈσ σ]	[σ σ ˈσ σ̆ σ σ]	[σ σ ˈσ σ̆ σ σ]	[σ σ ˈσ σ̆ σ σ]
/σ σ σ̆ σ̆ σ σ/	[σ σ σ̆ σ̆ ˈσ σ]	[σ σ σ̆ σ̆ ˈσ σ]	[σ ˈσ σ̆ σ̆ σ σ]	[σ ˈσ σ̆ σ̆ σ σ]
/σ σ̆ σ̆ σ̆ σ σ/	[σ σ̆ σ̆ σ̆ ˈσ σ]	[σ σ̆ σ̆ σ̆ ˈσ σ]	[σ σ̆ σ̆ σ̆ ˈσ σ]	[ˈσ σ̆ σ̆ σ̆ σ σ]
/σ σ̆ σ̆ σ̆ σ̆ σ/	[σ σ̆ σ̆ σ̆ σ̆ ˈσ]	[σ σ̆ σ̆ ˈσ̆ σ̆ σ]	[σ σ̆ σ̆ ˈσ̆ σ̆ σ]	[ˈσ σ̆ σ̆ σ̆ σ̆ σ]

- Learners trained using Robust Interpretive Parsing (RIP; Tesar and Smolensky 2000, Jarosz 2013) under Maximum Entropy Grammar updated with Stochastic Gradient Ascent (MaxEnt-SGA; Jäger 2007).
  Compare Staubs 2014, Hughto 2020, O'Hara 2021, and Sandell 2023 for similar learning frameworks.

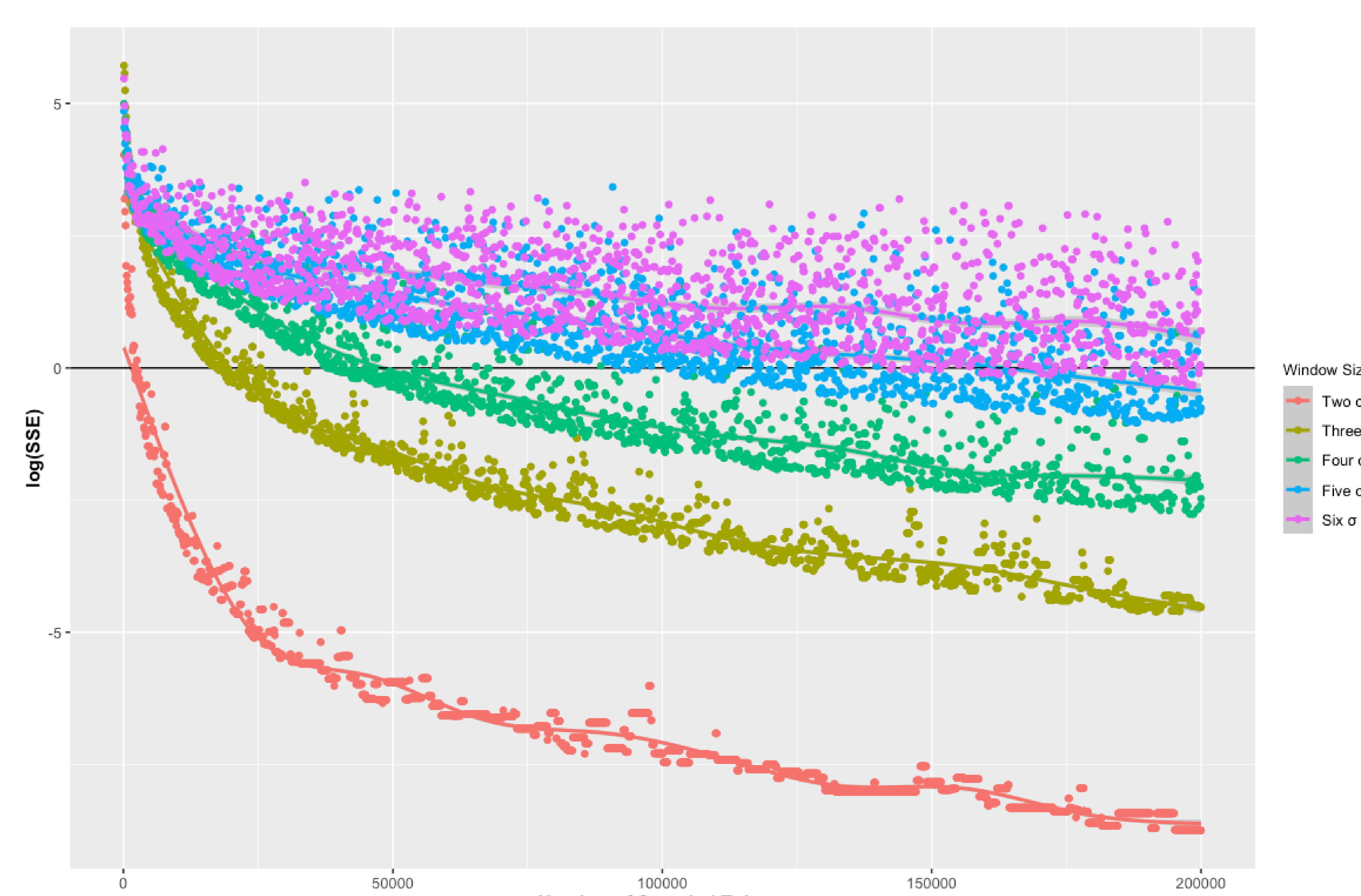
(6) Perceptron Update Rule (cf. Jarosz 2016)

$$w_{new} = w_{old} + \Delta w_i, \text{ where } \Delta w_i = \eta \times (L_i - W_i)$$

Token Dist.	**Evaluation**	**Convergence**	**Learning Rate**
Equiprobable	Sum squared error	SSE < 1	$\eta = 0.1$

- MARKEDNESS ($W = 5$) ≫ FAITHFULNESS ($W = 0$) in the initial state.
- 100 runs for each stress pattern; termination when convergence condition reached or 1000000 tokens sampled.

8. SSE Trajectory by Window Size



- Larger window size ⇒ less stable, overall slower decrease in SSE.
- Larger stress windows are more error-prone: RIPs more frequently cause adjustments to constraint weights that move the target grammar in the wrong direction overall (= credit problem; Dresher 1999).
- Calibrating constraint weights (here: ALL-Ϝ-R, ID-[stress], NONFIN-Ϝ) becomes more challenging as window sizes increase.
- Larger windows presumably yet more disadvantaged when the frequency distribution of types is not equiprobable (future simulations).
- ☞ Successfully learning large stress windows is possible, but requires correspondingly more data (and patience).

9. Simulation Results

Window Size	Mean G Updates	Mean Tokens
2	177	2346
3	1175	18058
4	1663	38942
5	3064	96512
6	3862	152880

Highly significant correlation between window size and the average number of grammar updates and average number of training tokens to convergence:

- Mean grammar updates: $r = 0.99, t = 13.525, p < 0.001$
- Mean training tokens: $r = 0.965, t = 6.363, p < 0.01$

10. Discussion

1. Staubs (2014: 94–98) claimed that four-syllable stress windows are learnable, and though very rare, not perhaps expected to be entirely absent from a typology of stress systems.
   - See now Sandell 2023: Ch. 5 for evidence for a four-syllable stress window in post-Vedic ("Classical") Sanskrit.
2. Stanton (2016) argued that the credit problem and the relatively lower frequency of longer words make stress systems exhibiting a species of midpoint pathology difficult to acquire.
3. Results presented here on Paamese-type window systems support the conclusion that **"soft" limits** on learnability are largely responsible for many aspects of the typology of stress systems.

References

Crowley, Terry. 1982. *The Paamese Language of Vanuatu*. No. 68 in Pacific Linguistics Series B. Canberra: Australian National University.; **Dresher**, Elan B. 1999. Charting the Learning Path: Cues to Parameter Setting. *Linguistic Inquiry* 30.27–67.; **Goedemans**, Rob, Jeffrey Heinz, and Harry van der Hulst. 2014. StressTyp2. URL: `http://st2.ullet.net`.; **Goldsmith**, John. 1990. *Autosegmental and Metrical Phonology*. Oxford: Basil Blackwell.; **Hayes**, Bruce. 1995. *Metrical Stress Theory*. Chicago: University of Chicago Press.; **Hughto**, Coral. 2020. Emergent Typological Effects of Agent-Based Learning Models in Maximum Entropy Grammar. Ph.D. diss., University Of Massachusetts, Amherst.; **Jäger**, Gerhard. 2007. Maximum Entropy Models and Stochastic Optimality Theory. In Jane Grimshaw, J. Maling, Christopher D. Manning Manning, J. Simpson and A. Zaenen (eds.), *Architectures, Rules, and Preferences: A Festschrift for Joan Bresnan*, 467–79. Stanford: CSLI Publications. **Jarosz**, Gaja. 2013. Learning with Hidden Structure in Optimality Theory and Harmonic Grammar: Beyond Robust Interpretive Parsing. *Phonology* 30.27–71.; **Jarosz**, Gaja. 2016. Investigating the Efficiency of Parsing Strategies for the Gradual Learning Algorithm. In Jeffrey Heinz, Rob Goedemans and Harry van der Hulst (eds.), *Dimensions of Phonological Stress*, 201–30. Cambridge: Cambridge University Press.; **Kager**, René. 1999. Optimality Theory. Cambridge: Cambridge University Press.; **Kager**, René. 2012. Stress in Windows: Language Typology and Factorial Typology. *Lingua* 122.1454–93.; **de Lacy**, Paul. 2020. The Feature [stress]. In Eno-Abasi Urua, Francis Egbokhare, Olúṣèyẹ Adéṣọlá and Harrison Adeniyi (eds.), *African Languages in Time and Space: A Festschrift in Honour of Professor Akinbiyi Akinlabi*, 1–27. Ibadan, Nigeria: Zenith BookHouse Ltd.; **Lee**, Minkyung. 1999. An Optimality Theoretic Analysis of Paamese Stress. In John Kyle (ed.), Kansas Working Papers in Linguistics, vol. 24, 59–80. Lawrence, KS: Linguistic Graduate Student Association, University of Kansas. URL: `https://kuscholarworks.ku.edu/bitstream/handle/1808/354/ling.wp.v24n1.paper4.pdf`.; **Legendre**, Géraldine, Antonella Sorace, and Paul Smolensky. 2006. The Optimality Theory-Harmonic Grammar Connection. In Paul Smolensky and Géraldine Legendre (eds.), *The Harmonic Mind. From Neural Computation to Optimality-Theoretic Grammar, Volume 2: Linguistic and Philosophical Implications*, 339–402. Cambridge, MA: MIT Press.; **O'Hara**, Charlie. 2021. Soft Biases in Phonology: Learnability Meets Grammar. Ph.D. diss., University of Southern California.; **Sandell**, Ryan. 2023. Towards a Dynamics of Prosodic Change: Corpus-Based and Computational Studies in the Synchronic and Diachronic Prosodic Phonology of Indic, Greek, and Germanic. Habilitationsschrift, Ludwig-Maximilians-Universität München.; **Stanton**, Juliet. 2016. Learnability Shapes Typolgy: The Case of the Midpoint Pathology. Language 92.753–91.; **Staubs**, Robert. 2014. Computational Modeling of Learning Biases in Stress Typology. Ph.D. diss., University Of Massachusetts, Amherst.; **Tesar**, Bruce, and Paul Smolensky. 2000. Learnability in Optimality Theory. Cambridge, MA: MIT Press.

GitHub

Data, code, and references available here: