Paamese-type Arbitrary-sized Stress Windows are Theoretically Learnable

(if you are very patient)

Ryan Sandell
Ludwig-Maximilians-Universität München

29.05.2025

1 Research Question

1. What is the relationship between **typology** and **learnability** in stress systems?
   - Are comparatively rarer and unattested stress patterns rare and/or absent from the attested typology because they are demonstrably **more difficult** or strictly **impossible** to learn?
2. Specifically: are STRESS WINDOWS subject to "hard" or "soft" limits on learnability?

(1) **Def.:** STRESS WINDOW
A distance in syllables (σ) from the left or right edge of a prosodic word (ω) within which primary stress is permitted to occur.

a. **Hard Limits:** Constraints / Parameters or similar furnished by UG strictly exclude conceivable patterns from occurring in natural languages.
   - Under parametric approaches to the modeling of stress systems (Halle and Vergnaud 1987, Dresher and Kaye 1990, Hayes 1995, Halle and Idsardi 1995, van der Hulst 2012; cf. Nazarov and Jarosz 2017), patterns that do not emerge from logically possible combinations of parameters are predicted to be nonexistent.
   - Accommodating novel empirical findings may require including additional parameter settings and/or constraints.
   - Reverse argument: patterns potentially generated by logically possible parameter settings or the factorial typology of constraint rankings that are unattested in the typology may indicate that the set of parameters or constraints employed is insufficiently restrictive.

b. **Soft Limits:** Certain stress patterns are theoretically learnable, but are disadvantaged relative to other patterns.
   - Disadvantage 1: relative rarity of disambiguating data.
   - Disadvantage 2: the credit problem (Dresher 1999) will more often produce grammar movements in the wrong direction when the same overt stress position might be attributed to opposing constraints.
     E.g., is stress on the leftmost syllable of a six-syllable word because it has a large right-edge window or a small left-edge window?
   - These combined disadvantages constitute the "long word problem" discussed in Stanton 2016.

3. **Claim:** there is no theoretical upper bound on the size of learnable stress windows.

    - Given sufficient data from words of sufficient length, right-edge-oriented window systems of up to six syllables (with unstressable final syllables) are shown through simulation to be learnable.
    - A strong positive correlation between number of grammar updates and window size obtains, but even the largest window systems tested successfully can converge on a categorical pattern.
    - This learnability result extends mathematically to progressively larger windows: arbitrarily large stress windows of the type tested will converge on the target pattern, given sufficient data and patience.

4. **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 further discussion in Goldsmith 1990: 215–16, Hayes 1995: 178–79, Lee 1999, and Kager 2012: 1466.

2. **Primary Stress in Paamese:**

    a. Primary stress (no secondary stress is reported) occurs by default on the penult (in disyllables) or on the antepenult (in longer words); see (2)a./b.

    b. Stress preferentially falls on the *preantepenult* when the antepenult contains an “unstressable” vowel, /V̌/; see (2)c./d.

    c. Stress may be assigned to the penult in a word if the word contains only three syllables or also the preantepenult is unstressable; see (2)d./e.

(2) **Paamese Stress Algorithm:** assign primary stress to the antepenult, unless the antepenult is lexically unstressable (V̌), else to the preantepenult, unless the preantepenult is lexically unstressable, else to the penult.

(3) **Paamese Stress Patterns**

a.	*í.nau*	‘I’	*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’				

3. Further aspects of Paamese stress patterns (not considered here):

    - Crowley (1982: 27) states that disyllables have initial stress unless the penult is unstressable, in which case they have exceptional final stress, though he provides no examples.
    - In the absence of data, I assume that final stress is strictly excluded in Paamese, although its existence would pose no issues for the analyses or simulations discussed below.
    - Morphological structure also plays a role in Paamese stress assignment, in that any vowel at the right edge of a morpheme is treated as unstressable, e.g., /matu-vaa/ ‘we (pcl. excl.) went’ is realized as *mátŭvaa*.
    - For present purposes, all such cases in which stress depends on morphological structure can be grouped together with morpheme-internal lexically unstressable vowels.

4. The **Paamese-type stress system** involving lexically unstressable elements — as opposed to stress-attracting elements — results in a pattern in which the default antepenultimate stress does not mark the farthest boundary of the stress window.

   a. Stress falls within a window at the right edge.

   b. When the default cannot apply, stress prefers to move *farther* from the edge that defines its window.

   c. Stress falls only as close to the word edge of its domain as a last resort.

5. Empirical question: Is the Paamese reduced four-syllable window learnable using a standard inventory of constraints on metrical stress?

   - Extension: Are yet larger stress windows exhibiting the same type of behavior likewise learnable, and if so, how does their relative ease of learnability compare?

3 Existing Analysis: Lee 1999

1. Lee (1999) shows that the data in (3) can be correctly generated by treating each prosodic word as containing a single binary, trochaic, quantity-insensitive foot near the right edge of the word.

   - Stress on underlying unstressable vowels, /V̆/, is categorically banned. Lee implements this ban with a constraint *[ˈə].
   - Feet aligned with the right edge of the word are strongly dispreferred; hence, penultimate stress occurs only as a last resort.

(4) **Constraint Ranking in Lee 1999:** FtBin, Trochee, *[ˈə] ≫ NonFin-Ft ≫ All-Ft-R ≫ Parse-σ

2. Preantepenultimate stress is a consequence of the avoidance of violations of FtBin, *[ˈə], and NonFinality-Ƒ, at the cost of a larger number of violations of All-Ƒ-R.

(5) /na-tăhosi/ → [(ˈna.tă).ho.si] I am good'

/na-tăhosi/	FtBin	Trochee	*[ˈə]	NonFin-Ƒ	All-Ƒ-R	Parse-σ
a. ☞ (ˈna.tă).ho.si					∗∗	∗∗
b. na.tă.(ˈho.si)				∗!		∗∗
c. na.(ˈtă.ho).si			∗!		∗	∗∗
d. na.(tă.ˈho).si		∗!			∗	∗∗
e. na.tă.(ˈho).si	∗!				∗	

3. Lee's constraint ranking will potentially result in **overgeneration:** the ranking in (4 yields an unbounded stress pattern, with primary stress on the rightmost stressable syllable that avoids a violation of NonFin-Ƒ.

   - A word of five or more syllables containing an unstressable antepenult and preantepenult, /σ_1 σ̆ σ̆ σ σ/, will receive primary stress on the stressable fifth-to-last syllable per (4).
   - The absence of an output like [ˈσ σ̆ σ̆ σ σ] in Paamese must by explained by the grammar, given Richness of the Base.

(6) Overgeneration: /σ σ̆ σ̆ σ σ/ → [(ˈσ σ̆) σ̆ σ σ]

/na-tăhosi/			FtBin	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			∗!		∗	∗∗∗

4. **On the constraint *[ˈə]**

- Although Lee (1999) expresses the avoidance of stress on particular syllable peaks through a markedness constraint, the unstressable vowels of do not possess any surface-true property that distinguishes them from stressable vowels.
- The substance of *[ˈə] is thus rather that of a faithfulness constraint: certain vowels are lexically marked in the input as stress-avoiding.
- This stress-avoiding behavior could be understood as an input specification [−stress] on certain vowels, in line with the interpretation of lexical stress proposed in de Lacy 2020.
- Thus, a faithfulness constraint **Ident-[stress]** is suitable for implementing the behavior of unstressable vowels in Paamese.
- Given that feet in Paamese are best treated as trochees, the same effect could be obtained by marking the unstressable vowels as foot tails (i.e., the right edge of a foot).
- See Özcelik 2014 and Yawney 2018 on Turkish, Yates 2020 and Sandell 2023: Ch. 4 on Vedic Sanskrit, using Anchor-L/$R_{Ϝ}$.

4 Window Restriction and Constraint Ganging

1. If the stress pattern of Paamese is not unbounded, how can stress be maintained with a window of four syllables at the right edge? Two OT solutions:

   a. Self-Conjunction (Ito and Mester 2003):

   - All-Ϝ-R/L constraints produce self-conjoined versions, e.g., All-Ϝ-R $\times 2$, All-Ϝ-R $\times 3$, etc., which can be ranked independently.
   - In the attested Paamese case, ranking All-Ϝ-R $\times 3 \gg$ Ident-[stress] would yield the desired effect.

   b. Endless *Lapse Edge (Gordon 2002: 503; cf. Elenbaas and Kager 1999, Kager 2012: 1478–9):

   - With only a single primary stress, general *Lapse and *Extended Lapse must be low-ranked in the grammar, since words of three or more syllables will frequently violate one or both constraints.
   - *(Extended)-Lapse-R/L are too restrictive for attested Paamese, since high-ranking *Extended-Lapse-R will exclude stress on the preantepenult.
   - To generate arbitrarily large windows, and arbitrarily large number of corresponding *Lapse-R/L constraint would need to be furnished.

2. In any formalism that permits constraint (self-)ganging (Harmonic Grammar, Maximum Entropy Grammar, etc.), ALL-Ƒ-L/R alone can transparently act as a window-size restrictor, without any additional assumptions.

    - See already Legendre et al. (2006) for this observation.
    - Given the constraints in (4), Linear OT (Potts et al. 2010) will find weights that allow for windows of arbitrary size $N + 2$, where N is the number of violations assigned to ALL-Ƒ-R in the leftmost permissible stress position
    - In the specific Paamese case, ruling out stress on the fifth-to-last syllable is straightforward: $3 \times W(\text{ALL-Ƒ-R}) > W(\text{NONFIN-Ƒ})$.

3. Allowing for stress windows of potentially arbitrary size brings an evident danger of overgeneration in terms of stress typology:

    - Four-syllable windows (full or reduced) are at best very rare, and possible cases are to some extent disputed.
        - Infrequent / morphologically restricted preantepenultimate stress in Standard Italian and Romanian (Roca 1999: 666–683)
        - Preantepenultimate stress due to considerations of vocalic prominence in Ashéninca (Payne 1990).
        - Preantepenultimate stress in right-edge moraic trochee systems: Old Latin (Allen 1973: 188–190, Jacobs 2003); Classical Arabic and some modern varieties (Watson 2011: §11; Janssens 1972); Classical Sanskrit (Sandell 2023).
    - Stress windows of size five or more are certainly unattested.

    ⇒ Assuming that stress windows are best explained via constraint ganging, why are larger stress windows generally, and stress windows of the **Paamese-type** specifically, nonetheless unattested?

(7) Frequency of stress window sizes in Goedemans and van der Hulst 2009 (see Kager 2012: 1464)

Window Size	Phonological	Lexical	Mixed	Total
Final two σ	59	20	3	82
Final three σ	22	12	4	38
Initial two σ	26	10	3	39
Initial three σ*	1	0	0	1
Total	108	42	10	160

5 Simulation Preparation

1. **Objective:** test the readiness with which categorical Paamese-like stress grammars of different window sizes are learned via exposure to overt forms marked only for primary stress (i.e., no “hidden” foot structure).

2. Data for these simulations consist of inputs (URs consisting of sequences of stressable and unstressable syllables) and candidates of overt forms marked for primary stress mapping to representations with a single foot parse.

    - Strings consist of stressable (σ) and unstressable (σ̆) syllables, 2–8 syllables in length.
    - Learners are assumed to know which syllables are lexically unstressable; these lexical properties are not learned simultaneously.

3. All inputs, overt forms, and parses were generated through scripting in R Core Development Team 2024.

(8) Data Generated for Simulations

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

4. Each mapping of an overt form mapping to a parse is automatically assessed for violations of 11 markedness constraints (see Kager 1999: Ch. 4) on metrical stress and the faithfulness constraint Ident-[stress].

(9) Constraint Inventory:
Align-L/R(ω, ƒ),
All-ƒ-L/R, FtBin, Trochee, Iamb, Parse-σ, NonFinality-ƒ, NonFinality-σ, NonInitiality-ƒ,
Ident-[stress].

- Violations of the further constraints MainRight, MainLeft, *Lapse-R/L, and *ExtendedLapse-R/L were also assessed, but these constraints were excluded in the simulations reported here.

6 Stress Patterns Evaluated

1. Under a given categorical stress pattern, for each input, there is exactly one matching overt form, though multiple possible parses that can yield the same overt form.
   - E.g., the summed probabilities of the parses [(ˈσ) σ] and [(ˈσ σ)] for the input /σ σ/ give the probability of the overt form [ˈσ σ] under a learner's grammar.
   - Thus, each of the 3584 overt forms has a specific probability (0 or 1), with only 508 having probability 1 under a given categorical stress pattern.
2. The expected probabilities for the overt forms corresponding to three quantity-insensitive right-edge fixed-stress systems and five right-edge window systems were generated.
   - QI right-edge fixed stress: final, penultimate, antepenultimate.
   - Right-edge window: 2–6 syllables.
   - The two- and three-syllable window patterns are "full" windows, permitting final stress; the 4–6-syllable window systems are "reduced", banning final stress.

(10) Expected overt outputs for six-syllable inputs

	Window Size				
Input	2	3	4 = Paamese	5	6
/σ σ σ σ σ σ/	[σ σ σ σ ˈσ σ]	[σ σ σ <u>**ˈσ**</u> 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 σ σ̆ <u>**ˈσ**</u>]	[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 σ σ̆ σ̆ <u>**ˈσ**</u>]	[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 σ σ̆ ˈσ σ]	[σ σ <u>**ˈσ**</u> 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 σ σ̆ σ̆ ˈσ σ]	[σ <u>**ˈσ**</u> 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]	[σ σ̆ σ̆ σ̆ ˈσ σ]	[<u>**ˈσ**</u> σ̆ σ̆ σ̆ σ σ]
/σ σ̆ σ̆ σ̆ σ̆ σ/	[σ σ̆ σ̆ σ̆ σ̆ ˈσ]	[σ σ̆ σ̆ σ̆ σ̆ ˈσ]	[σ σ̆ σ̆ ˈσ̆ σ̆ σ]	[σ σ̆ σ̆ ˈσ̆ σ̆ σ]	[ˈσ σ̆ σ̆ σ̆ σ̆ σ]

3. **Description of window systems**

    - Two-syllable window: default penultimate stress; final stress only if the penult is /σ̆/ and the final is /σ/.
    - Three-syllable window: default antepenultimate stress; penultimate stress only if antepenult is /σ̆/; final stress only if both penult and antepenult are /σ̆/.
    - 4–6-syllable windows: default antepenultimate stress; if the antepenult is /σ̆/, stress on the rightmost /σ/ other than the penult or final; stress on the penult only if all other syllables up to the 4th/5th/6th-to-last are /σ̆/; no final stress

7 Model Design and Parameters

1. Learners are 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.
    - Fundamental objective: optimize weights of constraints through exposure to overt forms without hidden prosodic structure.

2. **Maximum Entropy Grammar** (Goldwater and Johnson 2003, Jäger 2007, Hayes and Wilson 2008; cf. Jarosz 2019)

    - Each constraint has a real-valued weight, here assumed to be positive.
    - Each candidate, y in the set of candidates Y for a given input, x, has a Harmony, H, calculated as the sum of the number of violations of each constraint f_i multiplied by its corresponding weight w_i; see (11).
    - The probability of a candidate is the exponential of the negative harmony of y, divided by a normalization term Z.
    - The normalization term Z is the sum of the exponentiated negative harmonies of all candidates $y \in Y(x)$.

(11) Candidate Harmony

$$H(x, y) = \sum_{i=1}^{n} w_i f_i(y, x)$$

(12) Maximum Entropy Probability

$$P(y|x) = \frac{1}{Z} exp(-\sum_{i=1}^{n} w_i f_i(y, x))$$

(13) Maximum Entropy Normalization Term Z

$$Z = \sum_{y \in Y(x)} exp(-\sum_{i=1}^{n} w_i f_i(y, x))$$

- Any Harmonic Grammar capable of generating categorical outputs has a corresponding MaxEnt grammar.

3. **Robust Interpretive Parsing under MaxEnt**

   - Overt forms (e.g., [ˈσ σ̆ σ̆ σ]) are randomly sampled according to a frequently distribution.
   - Each overt form is associated to an input UR by stripping it of its overt stress ([ˈσ σ̆ σ̆ σ] ⇒ /ˈσ σ̆ σ̆ σ/); this is a TRAINING TOKEN.
   - For that input, the learner generates a PRODUCTION ORIENTED PARSE (POP) and a ROBUST INTERPRETIVE PARSE (RIP) using its current grammar.
   - Under a MaxEnt grammar, the POP is found by calculating the probabilities of all candidates given current constraint weights and sampling one candidate according to those probabilities.
   - The RIP is determined by identifying only the candidates leading to an overt form matching the overt form of the training token, and selecting one according to their MaxEnt probabilities.
   - If the overt forms of the POP and RIP match, no update to the grammar occurs.
   - If the overt forms of the POP and RIP differ, the grammar is updated to become more similar to the parse of the RIP. Grammar updates are thus error-driven.

4. **Stochastic Gradient Ascent**

   - The optimization of constraint weights proceeds by updating constraint weights using the Perceptron update rule (Rosenblatt 1958) when POP $\neq$ RIP.

   (14) Perceptron Update Rule (cf. Jarosz 2016)
   $w_{new} = w_{old} + \Delta w_i$, where $\Delta w_i = \eta \times (L_i - W_i)$

   - The POP is treated as the losing candidate, L_i, while the RIP is treated as the winning candidate W_i.
   - The difference in violation profiles between POP and RIP is multiplied by a learning rate, η.
   - Constraints violated by the RIP (winner) decrease in weight, while constraints violated by the POP (loser) increase in weight.

5. **Simulation Parameters**

   - In order to concentrate on the theoretically learnability of unattested windows for Paamese-type systems, all overt forms of all word lengths were treated as **equiprobable**.
   - The fit of a learner's MaxEnt grammar to a target categorical grammar is assessed by sum squared error (SSE): the square of the sum of the difference between the probability of each overt form under the learner's grammar and the probability of that overt form in the target grammar.
   - SSE is calculated after every 100 training tokens; whenever the grammar first obtains a SSE < 1, the grammar is considered to have converged on the target pattern.
   - A consistent rate $\eta = 0.1$ is used throughout.
   - All markedness constraints are given an initial weight of 5, thus preferring candidates with relatively fewer markedness violations in the initial state.
   - IDENT-[stress] is given an initial weight of 0.

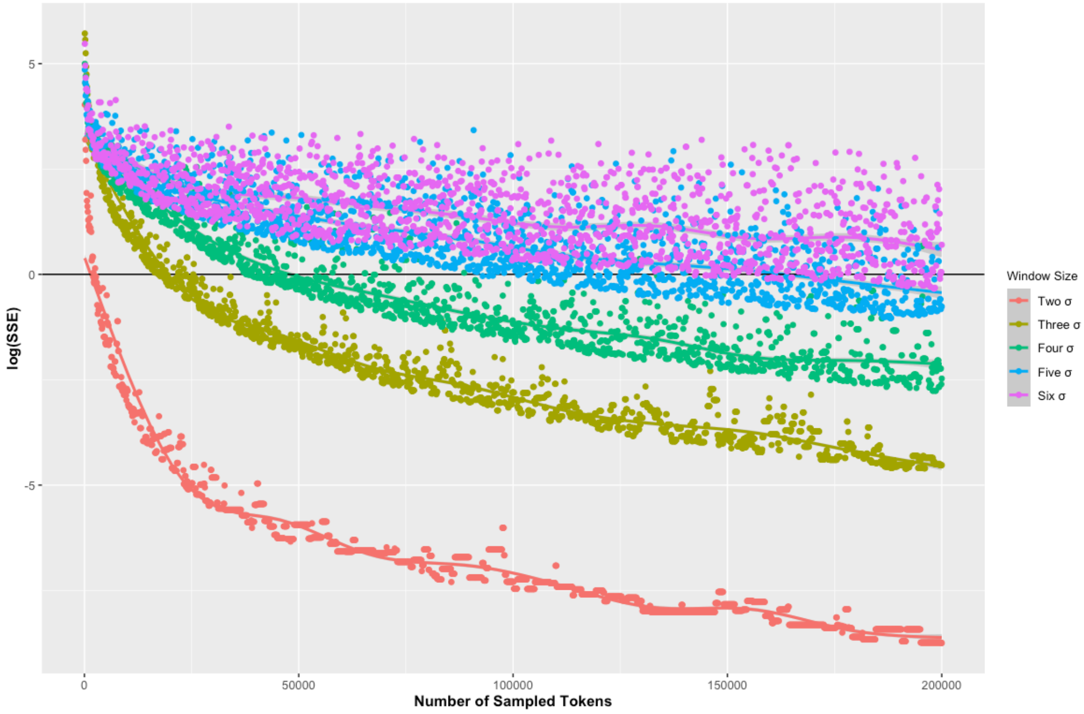
6. For each stress window pattern tested, 100 simulation runs were carried out.

   - A simulation run was terminated whenever the convergence condition SSE < 1 obtained.
   - In the worst case, the convergence condition was reached after exposure to ca. 170,000 tokens.

8 Results 1: SSE Trajectory by Window Size

1. (15) shows the gradual decrease in (log-transformed) SSE for the Paamese-type window systems.
   - In each of the simulations shown here, learning was terminated after exposure to 200,000 tokens.
   - Each point represents an SSE value at an interval of 100 tokens.

(15) Trajectory of SSE for 2–6-syllable Paamese-type Stress Grammars



2. Evident here is a relationship between a larger window size and a slower, less stable decrease in SSE.
   - This observation suggests that the larger windows are comparatively more difficult to learn.
   - Note that, for smaller windows, SSE mostly decreases or remains stable.
3. What causes intermediate increases in SSE?
   - Evidently, a Robust Interpretive Parse selected by the learner results can result in a change in constraint weights that alter the overall predictions of the grammar in a fashion that makes it less similar to the target pattern.
   - Consider the overt form [ˈσ σ̆ σ̆ σ̆ σ σ] provided to a learner acquiring a six-syllable window system.
     - If the learner selects a POP [σ σ̆ σ̆ (ˈσ̆ σ) σ] and a RIP [(ˈσ σ̆) σ̆ σ̆ σ σ], the result will be an increase in weight for ALL-Ƒ-R and IDENT-[stress], but a decrease in weight for ALL-Ƒ-L.
     - All candidates with stress closer to the left edge of the word will therefore gain in probability.
     - **Credit Problem:** the learner does not know that initial stress in [ˈσ σ̆ σ̆ σ̆ σ σ] should be attributed to IDENT-[stress] and NONFIN-Ƒ alone.
   - The learner must ultimately conclude that the correct parse, [(ˈσ σ̆) σ̆ σ̆ σ σ], is not due to any real effect of ALL-Ƒ-L or ALIGN-Wd-L, but rather due to high weights of IDENT-[stress] and NONFIN-Ƒ.
   - Conversely, when overt stress positions are limited to, e.g., the final and penultimate syllables, decreases in the weight of right-edge oriented markedness constraints are likely whenever error-driven learning occurs.

4. For any given stress window, only yet words longer than the size of the window are truly disambiguating.
   - For a six-syllable stress window, only words of seven syllables or more will provide an opportunity for the learner to determine that stress is not subject to any genuine left-edge attraction.
   - The decreased availability in terms of types of longer words for larger windows means that disambiguating data will be encountered less frequently.
   - Intuitively, if longer words exhibit lower token frequencies as well, the likelihood of having the opportunity to make useful grammar updates decreases yet further.

☞ Successfully learning constraint weights able to generate larger stress windows requires more grammar updates and exposure to more learning data.
   - Learning large windows with Paamese-type patterns naturally requires sufficient access to long words.
   - At the same time, sufficient instances of penultimate stress are required to ensure that the weight of NONFINALITY-Ƒ does not grow to large, and to help reduce the weight of left-edge oriented constraints.

9 Results 2: Correlations with Window Size

1. Mean values for the number of grammar updates and number of total training tokens sampled were calculated for each Paamese-type window system based on the 100 simulations run for each window pattern.

(16) Mean Grammar Updates and Training Tokens by Window Size

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

2. Pearson's correlation coefficient (called with `cor.test()` in R v. 4.4.2) and generalized linear models (called with `glm()` in R v. 4.4.2) show that a highly significant relationship between window size and number of grammar updates or number of training tokens to convergence exists.

(17) Mean grammar updates: $r = 0.99, t = 13.525, p < 0.001$

	Estimate	**Std. Error**	t	$p(> \|t\|)$
Intercept	-1721.5	290.78	-5.92	0.009629
Window Size	926.97	68.54	13.53	0.000874

(18) Mean training tokens: $r = 0.965, t = 6.363, p < 0.01$

	Estimate	**Std. Error**	t	$p(> \|t\|)$
Intercept	-90061	25307	-3.559	0.03785
Window Size	37952	5965	6.363	0.00786

- In both models, the t value found for the independent variable WINDOW SIZE is greater than the t value of the intercept.

3. From the generalized linear models, a corresponding increase in training data and number of grammar updates is predicted for yet larger window sizes.
    - E.g., for Paamese-type right-edge windows of sizes 7 and 8, 4767 and 5694 average grammar updates until convergence are predicted.
    - **Caution:** if the overt forms contain word forms only up to eight syllables in length, an 8-syllable window amounts to a default-opposite unbounded pattern, which may struggle to converge.
        - If yet longer word forms (9 or more syllables in length) are introduced into the data, the essential results are predicted to remain, but the specific quantitative relationship between window size and training data reported in (17) and (18) may be subject to change.

10 Discussion and Future Research

1. The results of the simulations conducted using RIP under MaxEnt-SGA reported in Sections 8 and 9 demonstrate that Paamese-type right-edge oriented window systems are theoretically learnable based on unstructured overt forms.
    - Intuitively and mathematically, this result extends to arbitrarily large windows, as long as correspondingly longer words are made available in the training data.
2. That stress systems with windows above the empirically well-established limit of 3 should be learnable accords with the finding of Staubs (2014: 88–98).
    - Employing a similar model design to that used in the simulations reported here (with slightly different constraints), Staubs carried out single-agent iterated learning simulations, found that windows larger than size 2 were subject to mislearning as opposite-edge systems.
    - In the iterated learning paradigm, windows larger than size 3 tend to simplify to three-syllable window systems, which in turn may be mislearned and two-syllable window systems over a series of generations.
    - Based on the probabilities of windows of given sizes resulting from the iterated learning simulations, Staubs finds that, given 160 window systems reported in Kager 2012, 1.6 languages would be expected to exhibit a four-syllable window.
    - Cases like Paamese and Classical Sanskrit (see §4.3 above) potentially fill this gap.
3. The empirical absence of stress windows of size five and above — whether of the Paamese-type (stress repelling), or involving lexical stress attraction or syllable weight — likely a function of the "long-word problem" as discussed in Stanton 2016.
    - When the likelihood of encountering longer words is relatively high — as in the simulations carried out here — relatively larger windows are indeed learnable.
    - Nonetheless, converging on the correct target grammar becomes increasingly difficult as the window size increases.
    - In a setting in which the probabilities of seven- and eight-syllable words are very small, learning five- and six-syllable windows is hypothesized to be prohibitively difficult.
4. **Conclusion:** the extreme rarity and absence of stress windows of size four and above is likely best understood as a function of **"soft" bias**.
    - Constraints of the Align family, plus constraint ganging, readily capture the attested range of patterns.

- Larger window systems most likely encounter nearly insuperable hurdles when learning occurs in a realistic setting: disambiguating data is too rare, and windows of even size 3 may be reanalyzed as opposite edge systems.
- The peculiarities of the Paamese-type stress system (i.e., stress-repelling elements) do not appear to significantly impact theoretical learnability.

5. **Future Simulations:**

- Empirical learnability: to what extent does learnability degrade as longer words become statistically rare in the training data?
- Paamese-type systems with unbounded stress: learning default-to-opposite (DTO) systems with standard metrical stress constraints requires multiple feet when multiple "special" elements occur in a string (Walker 1996, Baković 1998); stress-rejecting elements in Paamese-type systems that result in only a single foot per ω are predicted to make DTO patterns unlearnable (despite their relative computational simplicity; cf. Koser and Jardine 2020).
- Iterated learning: how does the stability of Paamese-type systems compare with the stability of window systems driven by surface markedness (Weight-to-Stress) or lexical stress-attracting elements?
  Compare Staubs 2014, O'Hara 2021, and Sandell 2023 for iterated learning paradigms in phonology.

References

Allen, W. Sidney. 1973. *Accent and Rhythm*. Cambridge: Cambridge University Press.

Baković, Eric. 1998. Unbounded Stress and Factorial Typology. In Ron Artstein and Madeline Holler (eds.), *RuLing Paper I: Working Papers from Rutgers University*, 15–28. New Brunswick, NJ: Department of Linguistics, Rutgers University. Available online as ROA-244, Rutgers Optimality Archive, http://roa.rutgers.edu/.

Crowley, Terry. 1982. *The Paamese Language of Vanuatu*. No. 68 in Pacific Linguistics Series B. Canberra: Australian National University.

Dresher, B. Elan, and Jonathan D. Kaye. 1990. A Computational Learning Model for Metrical Phonology. *Cognition* 34.137–95.

Dresher, Elan B. 1999. Charting the Learning Path: Cues to Parameter Setting. *Linguistic Inquiry* 30.27–67.

Elenbaas, Nine, and René Kager. 1999. Ternary Rhythm and the Laspe Constraint. *Phonology* 16.273–329.

Goedemans, Rob, and Harry van der Hulst. 2009. StressTyp: A Database for Word Accentual Patterns in the World's Languages. In M. Everaert, S. Musgrave and A. Dimitriadis (eds.), *The Use of Databases in Cross-Linguistic Studies*, 235–82. Berlin: Mouton de Gruyter.

Goldsmith, John. 1990. *Autosegmental and Metrical Phonology*. Oxford: Basil Blackwell.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In Jennifer Spenader, Anders Eriksson and Östen Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–20. Stockholm: Stockholm University Department of Linguistics.

Gordon, Matthew. 2002. A Factorial Typology of Quantity-Insensitive Stress. *Natural Language and Linguistic Theory* 20.491–552.

Halle, Morris, and William Idsardi. 1995. General Properties of Stress and Metrical Structure. In John A. Goldsmith (ed.), *The Handbook of Phonological Theory*, 403–443. Oxford: Blackwell.

Halle, Morris, and Jean-Roger Vergnaud. 1987. *An Essay on Stress*. Cambridge, MA: MIT Press.

Hayes, Bruce. 1995. *Metrical Stress Theory*. Chicago: University of Chicago Press.

Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* 39.379–440.

Hughto, Coral. 2020. Emergent Typological Effects of Agent-Based Learning Models in Maximum Entropy Grammar. Ph.D. diss., University Of Massachusetts, Amherst.

van der Hulst, Harry. 2012. Deconstructing Stress. *Lingua* 122.1494–521.

Ito, Junko, and Armin Mester. 2003. *Japanese Morphophonemics. Markedness and Word Structure*. Cambridge, MA: MIT Press.

Jacobs, Haike. 2003. The Emergence of Quantity-sensitivity in Latin: Secondary Stress, Iambic Shortening and Theoretical Implications for 'Mixed' Stress Systems. In D. Eric Holt (ed.), *Optimality Theory and Language Change*, 229–48. Dordrecht: Kluwer.

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.

Janssens, Gerard. 1972. *Stress in Arabic and Word Structure in the Modern Arabic Dialects*. Leuven: Peeters.

Jarosz, Gaja. 2013. Learning with Hidden Structure in Optimality Theory and Harmonic Grammar: Beyond Robust Interpretive Parsing. *Phonology* 30.27–71.

———. 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.

———. 2019. Computational Modeling of Phonological Learning. *Annual Review of Linguistics* 5.67–90.

Kager, René. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

———. 2012. Stress in Windows: Language Typology and Factorial Typology. *Lingua* 122.1454–93.

Koser, Nate, and Adam Jardine. 2020. The Computational Nature of Stress Assignment. In Hyunah Baek, Chikako Takahashi and Alex Hong-Lun Yeung (eds.), *Proceedings of the 2019 Annual Meeting on Phonology*, 1–11. Washington, D.C.: Linguistic Society of America.

de Lacy, Paul. 2020. The Feature [stress]. In Eno-Abasi Urua, Francis Egbokhare, Olúṣẹ̀yẹ 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.

Nazarov, Aleksei, and Gaja Jarosz. 2017. Learning Parametric Stress without Domain-Specific Mechanisms. In Karen Jesney, Charlie O'Hara, Caitlin Smith and Rachel Walker (eds.), *Proceedings of the 2016 Annual Meeting on Phonology*. Washington, D.C.: Linguistic Society of America. URL `https://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/4010`.

O'Hara, Charlie. 2021. Soft Biases in Phonology: Learnability Meets Grammar. Ph.D. diss., University of Southern California.

Özcelik, Öner. 2014. Prosodic Faithfulness to Foot Edges: The Case of Turkish Stress. *Phonology* 31.229–69.

Payne, Judith. 1990. Asheninca Stress Patterns. In Doris L. Payne (ed.), *Amazonian Linguistics: Studies in Lowland South American Languages*. Austin: University of Texas Press.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt, and Michael Becker. 2010. Harmonic Grammar with Linear Programming: From Linear Systems to Linguistic Typology. *Phonology* 27.1–41.

R Core Development Team. 2024. R Version 4.4.2 (2024-10-31) – "Pile of Leaves". URL `http://www.r-project.org/`.

Roca, Iggy. 1999. Stress in the Romance Languages. In Harry van der Hulst (ed.), *Word Prosodic Systems in the Languages of Europe*, 659–811. Berlin: Mouton de Gruyter.

Rosenblatt, Frank. 1958. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review* 65.386–408.

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.

Walker, Rachel. 1996. Prominence-Driven Stress. URL `http://roa.rutgers.edu/files/172-0197/172-0197-WALKER-0-0.PDF`, manuscript.

Watson, Janet C. E. 2011. Word Stress in Arabic. In Marc van Oostendorp, Elizabeth Hume and Keren Rice (eds.), *The Blackwell Companion to Phonology*, 2990–3019. Oxford: Wiley-Blackwell.

Yates, Anthony. 2020. Lexical Accents are Underlying Foot Edges: Evidence from Vedic Sanskrit. In Mariam Asatryan, Yixiao Song and Ayana Whitmal (eds.), *Proceedings of the 50th Annual Meeting of the North East Linguistic Society*, vol. 3, 255–68. Amherst, MA: GLSA.

Yawney, Heather L. 2018. On the Right to Be Faithful: The Irregular Stress of Turkish Adverbials with -En. *Toronto Working Papers in Linguistics* 40.1–16.