A characterization of the grammars that satisfy Hayes' Shifted Sigmoids Generalization

Keywords: phonological variation; MaxEnt grammars; Hayes' wug-shaped curves

- This talk offers an explicit formulation of the Shifted Sigmoids Generalization pioneered by Hayes and Zuraw (2017) and Hayes (2021) and a complete characterization of the harmony-based probabilistic grammars that satisfy it, showing that they are essentially MaxEnt grammars.

- **Nasal substitution** (NS) in Tagalog (Zuraw 2010) coalesces the nasal at the end of an affix and the obstruent at the beginning of a stem into a single consonant that is nasal as the former and homorganic to the latter. For instance, /maŋ+bigáj/ ('to distribute') is realized as [mamigáj]. Whether an affix+stem concatenation undergoes NS in Tagalog cannot be predicted based on the identity of the affix and the quality of the stem-initial consonant. Yet, we can count the **empirical frequency** of NS for concatenations of a certain affix with stems starting with a certain obstruent. To illustrate, for the affixes /maŋ/ and /paŋ/ and for stems that start with /k/ and /b/, we obtain the empirical frequencies of NS in fig. 1. Is there anything special about these numbers?

	/maŋ/	/paŋ/
/k/	0.993	0.909
/b/	0.916	0.434

Figure 1

- The curve in figure 2a plots the **sigmoid** $\mathbf{S}(x) = \frac{1}{1+\exp(-x)}$. Since the sigmoid takes values between 0 and 1, we can plot frequencies onto it. Thus, fig. 2a plots onto the sigmoid the frequencies 0.993 and 0.916 of NS for the underlying concatenations with the affix /maŋ/. **Shifted sigmoids** $\mathbf{S}(\Delta + x) = \frac{1}{1+\exp(\Delta - x)}$ are obtained by adding a constant Δ to the argument of the sigmoid. Fig. 2b determines Δ so that the frequency 0.909 of NS for /paŋ+k/ falls on the shifted sigmoid right underneath the frequency for /maŋ+k/ with the same stem-initial stop. Hayes and Zuraw (2017) make the surprising observation that the frequency 0.434 of NS for /paŋ+b/ falls almost perfectly on the same shifted sigmoid right underneath the frequency of NS for /maŋ+b/ with the same stem-initial stop, as in fig. 2c. If we know three of the frequencies, we can predict the fourth!

- Hayes (2021) indeed shows that, for a variety of processes in a variety of languages (vowel harmony in Hungarian; liaison in French; final devoicing in Dutch; genitive plurals in Finnish; schwa/zero alternations in French; stress placement in Hupa), the empirical frequencies of the process applying on two-by-two underlying forms fall on two shifted sigmoids. What is driving this pattern? Let us focus on NS. Some constraints are sensitive to the identity of the affix but not to the quality of the stem-initial obstruent: the number of violations is constant across stem-initial obstruents, namely $C(/\text{maŋ+k}/, \text{NS}) = C(/\text{maŋ+b}/, \text{NS})$. Some other constraints are sensitive to the quality of the stem-initial obstruent but not to the identity of the affix: the number of violations is constant across affixes, namely $C(/\text{maŋ+k}/, \text{NS}) = C(/\text{paŋ+k}/, \text{NS})$. Crucially, no constraint is sensitive to both affixes <u>and</u> stems, because affixes and stems are **independent** phonological dimensions. Equivalently, every constraint is constant either across stem-initial obstruents or across affixes.

- These considerations lead to the following formulation of Hayes' **Shifted Sigmoids Generalization** (SSG). Suppose that four underlying forms can be organized into a two-by-two square along two phonological dimensions, as in fig. 3a. Suppose that these two phonological dimensions are independent in the sense that no relevant constraint is sensitive to both dimensions. Equivalently, every constraint is constant along one of the two dimensions, as stated in fig. 3b. Then, the empirical frequencies of the relevant process applying to those four forms can be fitted on two shifted sigmoids, as in fig. 3c. Once again, if we know three of the frequencies, we can predict the fourth.

- Which probabilistic phonological grammars satisfy this SSG? We focus on **harmony-based** grammars defined as follows. We start from a set $\mathbf{C}$ of n **constraints** $C_1, \ldots, C_n$ that assign to each

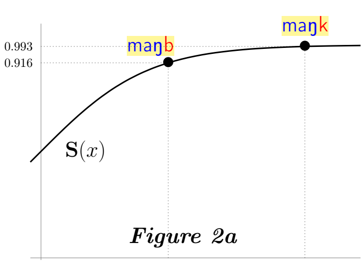

Figure 2a

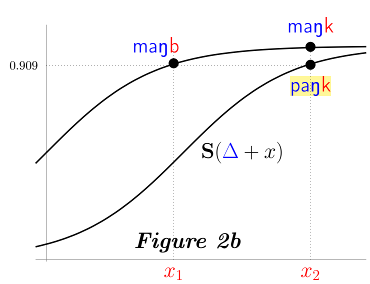

Figure 2b

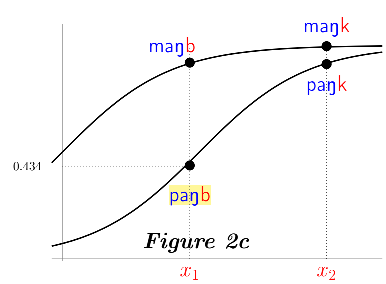

Figure 2c

Proceedings of the Society for Computation in Linguistics (SCiL) 2023, pages 418-419.
Amherst, Massachusetts, June 15-17, 2023

phonological mapping (UR, SR) an n-dimensional vector $\mathbf{C}(\mathsf{UR},\mathsf{SR}) = (C_1(\mathsf{UR},\mathsf{SR}), \ldots, C_n(\mathsf{UR},\mathsf{SR}))$ of constraints violations. We avail ourselves of a **harmony** H that assigns a non-negative harmony score $H(\boldsymbol{x})$ to any (integral, non-negative) n-dimensional vector $\boldsymbol{x} = (x_1, \ldots, x_n)$. We then define the probabilistic phonological grammar G_H as follows: the probability $G_H(\mathsf{SR} \,|\, \mathsf{UR})$ that a certain UR is realized as a certain candidate SR is the harmony score assigned by H to the vector of constraint violations of the mapping (UR, SR) divided by a constant Z that ensures normalization, namely $G_H(\mathsf{SR} \,|\, \mathsf{UR}) = \frac{1}{Z} H(\mathbf{C}(\mathsf{UR},\mathsf{SR}))$. To illustrate, when the harmony score $H(\boldsymbol{x})$ is defined through some non-negative weights w_k as $H(\boldsymbol{x}) = \exp(-\sum_{k=1}^{n} w_k x_k)$, the harmony-based probabilistic grammar G_H is a **MaxEnt grammar**.

- The main result of the talk is the following complete (both necessary and sufficient) characterization of harmony-based grammars that satisfy Hayes' SSG:

> *A harmony function H for n-dimensional vectors yields a harmony-based probabilistic grammar G_H that satisfies Hayes' SSG if and only if it is **separable** in the sense that there exist some functions $f_1, \ldots, f_n$ such that the harmony score $H(\boldsymbol{x})$ of any n-dimensional (integral and non-negative) vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ admits the expression $H(\boldsymbol{x}) = \prod_{k=1}^{n} f_k(x_k)$.*

Intuitively, separability $H(\boldsymbol{x}) = \prod_{k=1}^{n} f_k(x_k)$ says that the harmony H is simple because the decision of which score $H(\boldsymbol{x})$ to assign to a vector $\boldsymbol{x}$ can be broken up into decisions f_k each of which narrowly takes into account only one component x_k of the vector but ignores all other components (equivalently, the logarithm $\log H(\boldsymbol{x}) = \sum_{k=1}^{n} \log f_k(x_k)$ of the harmony H is a separable utility function in the sense of mathematical economics; Debreu 1960; Wakker 1988). Thus, this boxed result intuitively says that a constraint set that is simple (in the sense that it treats two phonological dimensions as independent) yields a pattern of probabilities that is simple (in the sense that any probability can be predicted from the other three probabilities because they sit on two shifted sigmoids) whenever the model of constraint interaction is simple (in the sense that it is based on a harmony that is simple because separable). The proof that a harmony of the form $H(\boldsymbol{x}) = \prod_{k=1}^{n} f_k(x_k)$ yields a grammar G_H that satisfies the SSG is straightforward. The proof of the reverse is non-trivial because it requires constructing the functions f_k starting from the SSG.

- The MaxEnt harmony $H(\boldsymbol{x}) = \exp(-\sum_{k=1}^{n} w_k x_k)$ recalled above satisfies the separability condition $H(\boldsymbol{x}) = \prod_{k=1}^{n} f_k(x_k)$ with $f_k(x) = \exp(-w_k x)$. Furthermore, let us suppose that candidate sets are infinite because they consist of all strings of finite but arbitrary length constructed out of a finite alphabet. A constraint C **grows at most linearly** provided the number of violations it assigns to a candidate surface string of length ℓ is never larger than $A\ell + B$, for some constants $A, B \geq 0$. To illustrate, the number of epenthetic segments is upper bounded by the number ℓ of surface segments. The constraint $C =$ Dep thus grows at most linearly, because the number of violations it assigns to a candidate of length ℓ is never larger than $A\ell + B$ with $A = 1$ and $B = 0$. Most constraint sets used in the literature consist of constraints that all grow at most linearly with the length of the candidate surface strings. In this case, we can reason as in Daland (2015) to show that the normalization constant Z for a separable harmony $H(\boldsymbol{x}) = \prod_{k=1}^{n} f_k(x_k)$ is finite only if every function f_k comes with some value w_k such that $f_k(x) \leq \exp(-w_k x)$ (definitely for large x). In other words, MaxEnt harmonies are the largest harmonies that satisfy Hayes' SSG.

	second dimension	
first	/form$_{1,1}$/	/form$_{1,2}$/
dimension	/form$_{2,1}$/	/form$_{2,2}$/

Figure 3a

C is constant wrt first dimension:

$C(/\mathsf{form}_{1,1}/, \ldots) \quad C(\mathsf{form}_{1,2}, \ldots)$

$\quad \| \qquad\qquad\qquad \|$

$C(/\mathsf{form}_{2,1}/, \ldots) \quad C(\mathsf{form}_{2,2}, \ldots)$

or C is constant wrt second dimension:

$C(/\mathsf{form}_{1,1}/, \ldots) = C(/\mathsf{form}_{1,2}/, \ldots)$

$C(/\mathsf{form}_{2,1}/, \ldots) = C(/\mathsf{form}_{2,2}/, \ldots)$

Figure 3b

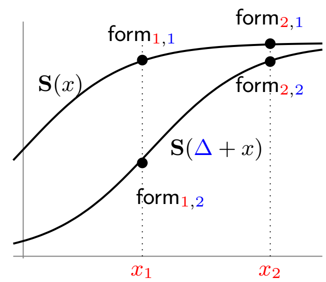

Figure 3c