Phonology **33** (2016). Supplementary materials

Relating application frequency to morphological structure: the case of Tommo So vowel harmony

Laura McPherson
Dartmouth College

Bruce Hayes
University of California, Los Angeles

Supplementary materials

The data corpus and model calculations can be found in the following Excel files:

- data corpus
- calculations for the maxent model
- calculations for the domain indexation model
- calculations for the exponential model
- calculations for the morpheme indexation model
- calculations for the noisy harmonic grammar model

The following text files are called in the Monte Carlo simulation in Appendix C: §4:

- Monte Carlo simulation data
- Monte Carlo simulation results

Appendix A: The math

1 Background and starting point

1.1 Goal of this document

- Explain and justify all of the math in our article.
- Intended audience: people who vaguely remember their math training from long ago.
- Why not in the article itself?
  - Takes a huge amount of space.
  - This isn't research, just standard math.
  - We couldn't find a textbook source that covers this completely or clearly.
  - By writing this out ourselves we can show every single step – much as one would in a traditional phonological derivation.
- Caveat to people who do math all the time:
  - The degree of detail included here is likely to be irritating! We recommend you stop reading after §2.3 and work out the results yourself as an exercise.

1.2 Substantive points to be established

1.2.1 *Our system of constraints derives sigmoids*

- If you apply the principles of maxent grammar to the system we described, with one scalar markedness constraint and one non-scalar faithfulness constraint, then you predict a sigmoid curve, described by the equation:

$$y = \frac{1}{1 + e^{F - Mx}}$$

y = predicted application rate of phonological process
F = weight of faithfulness
M = weight of applicable scalar markedness constraint
x = value on scale (for us: 'distance' from root along the morphological scale we define); higher values = closer

1.2.2 *Properties of the sigmoid function*

- It asymptotes at 1 for large values of x.
- It asymptotes at 0 for small values of x.
- It crosses 50% probability at x = M/F.
- It is symmetrical about this point.
- The maximum slope occurs at this point and is equal to M/4.

1.2.3 *A schematic sigmoid showing all these properties*

- Following the article, we chose a scale for x ranging from 1 to 7.
- For Fig. 1 below, we picked F = 8 and M = 2.
- So, F/M = 4, which is where the sigmoid curve crosses 0.5 (vertical blue line shows this).
- Symmetry about this point, and asymptotes at zero and one, are visually evident.
- The diagonal red line is the tangent to the sigmoid at the point of maximum slope. The slope of this line can be seen to be 0.5 (line rises by 1 in an interval of 2), which is equal to M/4, i.e. 2/4.

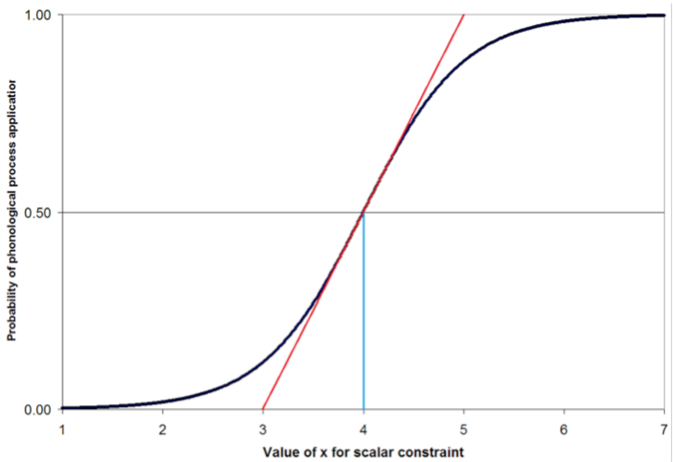

Figure 1

2 Math you need to remember

2.1 Algebraic identities

a. Multiplying exponentiated numbers is the same as adding their exponents

$a^x \times a^y = a^{(x+y)}$

b. Anything to the zeroth power is one.

$a^0 = 1$

c. Distribution of multiplication over addition

$a \times (b + c) = (a \times b) + (a \times c)$

d. Product of two sums

We can invoke the distributive property repeatedly to get this:

$$
\begin{aligned}
(a + b)(c + d) &= ((a + b) \times c) + ((a + b) \times d) \\
&= (c \times (a + b)) + (d \times (a + b)) && \text{(commute)} \\
&= ca + cb + da + db && \text{(distribute again, twice)} \\
&= ac + bc + ad + bd && \text{(commute)}
\end{aligned}
$$

3 Our system of constraints derives sigmoids

3.1 Starting point

- Let us assume a generic phonological process; in informal terms we think of the process as 'applying' or 'not applying.'
- We assume that the two essential conflicting constraints are a scalar markedness constraint and an opposing non-scalar faithfulness constraint.

3.2 Variables

- Let the weight of the scalar markedness constraint be M.
- Let the weight of the non-scalar faithfulness constraint be F.
- Let there be a variable x expressing the relevant scale. In our article this is root-closeness, so
  - 7 = root-internal
  - 1 = the farthest away affix.

3.3 Applying maxent

- Assume we're trying to decide the phonological outcome for an input whose value along the scale is x.

3.3.1 *Calculate harmony*

- The candidate that has undergone the phonological process violates (just) FAITH, so its harmony is F.
- The candidate that has not undergone the phonological process violates (just) the scalar markedness constraint, so its harmony is Mx.
- We assume that all other candidates are ruled out by very strong constraints – super-high weights, and so need not be taken into account.

3.3.2 *Negate and exponentiate*

- As our article notes, the next step in maxent is to take e to the power of the negative harmony, for each candidate.

  Candidate that has undergone phonology: $e^{-\mathrm{F}}$

  Candidate that has not undergone phonology: $e^{-\mathrm{M}x}$

  All other candidates: $e^{-(\text{some very large value})}$

  - e to some very large value is close enough to zero that we can justifiably treat these candidates as receiving the value zero.

3.3.3 *Calculate the denominator* ('Z')

- Z is the sum of what you got for all candidates.
- As noted above, most of these get vanishing small values, which we will ignore.
- So really, we just have two candidates to worry about, and their values sum to:

$$\mathrm{Z} = e^{-\mathrm{F}} + e^{-\mathrm{M}x}$$

3.3.4 *Find the probabilities of the candidates*

- In maxent, this is the result of step 3.3.2 above, divided by Z.

  a. *Probability of 'undergoing' candidate* $= \dfrac{\mathrm{e}^{-\mathrm{F}}}{\mathrm{Z}}$

  b. *Probability of 'non-undergoing' candidate* $= \dfrac{\mathrm{e}^{-\mathrm{M}x}}{\mathrm{Z}}$

3.4 Deriving the sigmoid curve

- We want to see how the probability of the 'undergoing' candidate varies with x, its value along the scale.

3.4.1 *The algebra*

- Start with the probability that we just derived for the probability of the 'undergoing' candidate (3.3.4a):

$$\textit{Probability of undergoing candidate} = \frac{\mathrm{e}^{-\mathrm{F}}}{\mathrm{Z}}$$

- Substitute in the formula for Z (3.3.3):

$$\textit{Probability of undergoing candidate} = \frac{\mathrm{e}^{-\mathrm{F}}}{\mathrm{e}^{-\mathrm{F}} + \mathrm{e}^{-\mathrm{M}x}}$$

- Multiply top and bottom by e^{F}:

$$\textit{Probability of undergoing candidate} = \frac{\mathrm{e}^{\mathrm{F}} \times \mathrm{e}^{-\mathrm{F}}}{\mathrm{e}^{\mathrm{F}} \times (\mathrm{e}^{-\mathrm{F}} + \mathrm{e}^{-\mathrm{M}x})}$$

- On the bottom, distribute multiplication over addition (2.1c):

$$\textit{Probability of undergoing candidate} = \frac{e^{F} \times e^{-F}}{(e^{F} \times e^{-F}) + (e^{F} \times e^{-Mx})}$$

- Multiplication of exponentiated terms is the same as exponentiation by the sum of the exponents (2.1a).

$$\textit{Probability of undergoing candidate} = \frac{e^{(F + -F)}}{e^{(F + -F)} + e^{F - Mx}}$$

- Summing:

$$\textit{Probability of undergoing candidate} = \frac{e^{(0)}}{e^{(0)} + e^{F - Mx}}$$

- Any positive number to the zeroth power is 1 (2.1b):

$$\textit{Probability of undergoing candidate} = \frac{1}{1 + e^{F - Mx}}$$

3.4.2 *Result*

- We are done! What we have is a version of the standard LOGISTIC FUNCTION, which plots as a sigmoid (see 1.2.3 above).
- The function derives a probability of the phonology-undergoing candidate from its value on the markedness scale, x.
- Next, let us provide mathematical demonstrations of the properties of this function laid out in 1.2.2 above.

4 First two properties of the sigmoid: limits as *x* is made high or low

4.1 Asymptotes to 1 as *x* becomes large

- We consider our sigmoid function, using y to mean 'probability of undergoing candidate':

$$y = \frac{1}{1 + e^{F - Mx}}$$

- The informal reasoning behind the asymptotes:
  - If x is very large, then Mx is very large too (we use positive weights).
  - Then $F - Mx$ is very large and negative.
  - Then $e^{F - Mx}$ is very small.
  - Then $\frac{1}{1 + e^{F - Mx}}$ approaches 1.
- Compare Tommo So, with near-obligatory root harmony.

4.2 Asymptotes to zero as x gets low

- Here, by 'small' we will include even negative numbers.
- Reasoning:
  - If x is very negative, then Mx is very negative too (we use positive weights).
  - Then F – Mx is very large and positive.
  - Then $e^{F - Mx}$ is very large.
  - Then $\frac{1}{1 + e^{F - Mx}}$ approaches zero.
- Compare Tommo So, with zero harmony in the 'outermost' inflectional levels.

5 Second property of the sigmoid: crosses 50% at $x = F/M$

5.1 Demonstration

- We plug the value F/M into our formula for the sigmoid, replacing x:

$$y = \frac{1}{1 + e^{F - Mx}}$$

$$= \frac{1}{1 + e^{F - (M \times F/M)}}$$

- The M terms cancel each other out and can therefore be removed:

$$= \frac{1}{1 + e^{(F - F)}}$$

- The F terms disappear:

$$= \frac{1}{1 + e^{0}}$$

- Anything to the zero is one (2.1b):

$$= \frac{1}{1 + 1}$$

- And so:

$$y = \frac{1}{2}$$

- This is 50%, the probability of undergoing phonology at the point F/M.

6 Third property of the sigmoid: symmetrical about $x = \mathrm{F/M}$

6.1 What we want to establish

- We know that at $x = \mathrm{F/M}$ the value of the function is 0.5 (see 5.1).
- Let us consider two points with values of x symmetric about F/M. Call them $x + \Delta x$ and $x - \Delta x$.
- The amount that the function computed at $x + \Delta x$ *exceeds 0*.5 should be the same as the amount that the function at point $x - \Delta x$ *falls below* 0.5. This is what is meant here by symmetry.

6.2 The algebra that demonstrates symmetry

- Start with an equality that looks arbitrary but is chosen with forethought and is obviously true:

  $1 + e^{M\Delta x} + e^{-M\Delta x} + 1 = 1 + e^{M\Delta x} + e^{-M\Delta x} + 1$

- Replace the last instance of 1 with its equal, e^0 (2.1b).

  $1 + e^{M\Delta x} + e^{-M\Delta x} + 1 = 1 + e^{M\Delta x} + e^{-M\Delta x} + e^0$

- Since $M\Delta x + -M\Delta x = 0$, we can replace the zero with this expression:

  $1 + e^{M\Delta x} + e^{-M\Delta x} + 1 = 1 + e^{M\Delta x} + e^{-M\Delta x} + e^{M\Delta x + -M\Delta x}$

- Since $e^{a+b} = e^a \times e^b$ (2.1a), we can rewrite the $e^{M\Delta x + -M\Delta x}$ term like this:

  $1 + e^{M\Delta x} + e^{-M\Delta x} + 1 = 1 + e^{M\Delta x} + e^{-M\Delta x} + e^{M\Delta x} X e^{-M\Delta x}$

- Since $(a + b)(c + d) = ac + ad + bc + bd$ (see 2.1d above), we can rewrite the right side like this:

  $1 + e^{M\Delta x} + e^{-M\Delta x} + 1 = (1 + e^{M\Delta x}) \times (1 + e^{-M\Delta x})$

  To check this, observe that the right side of the new version, when multiplied out, yields the right side of the old version.

- Divide both sides by $(1 + e^{M\Delta x})$:

  $$\frac{1 + e^{M\Delta x} + e^{-M\Delta x} + 1}{1 + e^{M\Delta x}} = \frac{(1 + e^{M\Delta x}) \times (1 + e^{-M\Delta x})}{1 + e^{M\Delta x}}$$

- Simplify on the right side by dividing top and bottom by $1 + e^{M\Delta x}$:

  $$\frac{1 + e^{M\Delta x} + e^{-M\Delta x} + 1}{1 + e^{M\Delta x}} = 1 + e^{-M\Delta x}$$

- Split up the left side, giving each term the same denominator:

  $$\frac{1 + e^{M\Delta x}}{1 + e^{M\Delta x}} + \frac{e^{-M\Delta x} + 1}{1 + e^{M\Delta x}} = 1 + e^{-M\Delta x}$$

- Simplify the fraction on the left side:

$$1+\frac{e^{-M\Delta x}+1}{1+e^{M\Delta x}}=1+e^{-M\Delta x}$$

- Divide both sides by $1+e^{-M\Delta x}$:

$$\frac{1}{1+e^{-M\Delta x}}+\frac{1}{1+e^{M\Delta x}}=1$$

- Subtract 0.5 from both sides:

$$\frac{1}{1+e^{-M\Delta x}}+\frac{1}{1+e^{M\Delta x}}-0.5=0.5$$

- Subtract $1/1+e^{-M\Delta x}$ from both sides:

$$\frac{1}{1+e^{M\Delta x}}-0.5=0.5-\frac{1}{1+e^{-M\Delta x}}$$

- Zero may be added without change:

$$\frac{1}{1+e^{M\Delta x+0}}-0.5=0.5-\frac{1}{1+e^{-M\Delta x+0}}$$

- Replace zero by $F-F$:

$$\frac{1}{1+e^{M\Delta x+F-F}}-0.5=0.5-\frac{1}{1+e^{-M\Delta x+F-F}}$$

- Rearranging order of addends in the exponent on both sides:

$$\frac{1}{1+e^{-F+M\Delta x+F}}-0.5=0.5-\frac{1}{1+e^{-F-M\Delta x+F}}$$

- Substitute in $M\times -F/M$ for $-F$ (it's the same thing, since the multiplication and division cancel each other out):

$$\frac{1}{1+e^{(M\times -F/M)+M\Delta x+F}}-0.5=0.5-\frac{1}{1+e^{(M\times -F/M)-M\Delta x+F}}$$

- Extract the common factor M (2.1c):

$$\frac{1}{1+e^{M\times(-F/M+\Delta x)+F}}-0.5=0.5-\frac{1}{1+e^{M\times(-F/M-\Delta x)+F}}$$

- Put negative signs on the outside:

$$\frac{1}{1+e^{-M\times(F/M-\Delta x)+F}}-0.5=0.5-\frac{1}{1+e^{-M\times(F/M+\Delta x)+F}}$$

- Swap order:

$$\frac{1}{1+e^{F-M\times(F/M-\Delta x)}}-0.5=0.5-\frac{1}{1+e^{F-M\times(F/M+\Delta x)}}$$

- This is just what we want. Recall that the formula deriving predicted probability of phonological application is:

  $$\frac{1}{1 + e^{F - M \times x}}$$

  On the left side, we evaluate the function for the value $x = F/M - \Delta x$, which is Δx below the 0.5 crossover value of F/M; and we compute how much this is above the crossover value of 0.5. On the right side, we evaluate the function for the value $x = F/M + \Delta x$, which is Δx *above* the crossover value; and we compute how much this is *below* the crossover value of 0.5. Since these turn out always to be equal, the function is symmetrical about the value F/M.

7 Fourth property of the sigmoids: slope at F/M = M/4

7.1 Finding the slope at the symmetry point

- In calculus the steepness of a curve at any given point is expressed as its DERIVATIVE.
- Figure 2 is the logistic curve plotted earlier as Fig. 1, this time plotted together with its derivative. The derivative has a symmetrical hump peaking at M/F (which you will remember is $8/2 = 4$), where (as we expected) it has the value 0.5, since this is M/4.

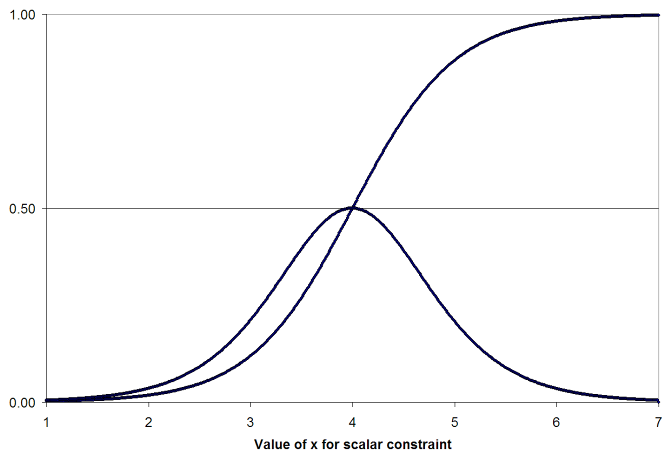

Figure 2

- Calculating the derivative of

  $$y = \frac{1}{1 + e^{F - Mx}}$$

  (see §9 below for how this can be done), it turns out to be:

  $$\frac{dy}{dx} = \frac{Me^{F-Mx}}{(1 + e^{F-Mx})^2}$$

  This is indeed the function that was used in plotting the derivative curve in the figure above.

- We can now profitably ask what the slope is at the symmetry point, which it will be recalled, is at $x = F/M$. We substitute F/M into the formula for the derivative, and obtain:

  $$\begin{aligned} \frac{dy}{dx}(F/M) &= \frac{Me^{F - M \times F/M}}{(1 + e^{F - M \times F/M})^2} \\ &= \frac{Me^{F - F}}{(1 + e^{F - F})^2} \\ &= \frac{Me^{0}}{(1 + e^{0})^2} \\ &= \frac{M \times 1}{(1 + 1)^2} \\ &= \frac{M}{(2)^2} \\ &= \frac{M}{4} \end{aligned}$$

- So, we've established that the slope at the symmetry point is the markedness weight divided by 4.

8 Fifth property of the sigmoids: M/4 is the maximum slope

8.1 The last detail

- We've found that the slope at F/M is indeed M/4, but we haven't shown that that is the *steepest* slope of the logistic curve.
- In calculus, the way to find a maximum of a function is to find the spot where its derivative levels out. For example, if we look again at (a pristine version of) the derivative of the logistic function that we plotted earlier, it appears like this:

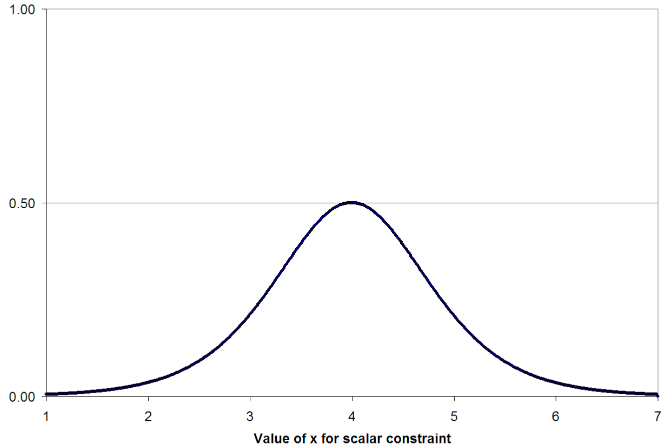

Figure 3

It would appear (and we'll nail this down shortly) that the maximum is indeed at $x = 4$, with the value 0.5. To verify this, we would want to know that a line tangent to the derivative function at this point would be level; i.e. would have slope zero. (Conveniently, the vertical scale line at $y = 0.5$ actually happens to be this tangent line.)

- So what we do is compute the SECOND DERIVATIVE; i.e. the derivative of the derivative, and find where it goes to zero.

8.2 The second derivative

- We'll examine the equation for the second derivative in a moment. But first, it would be useful just to look at the graph for it (plotted along with the first derivative):

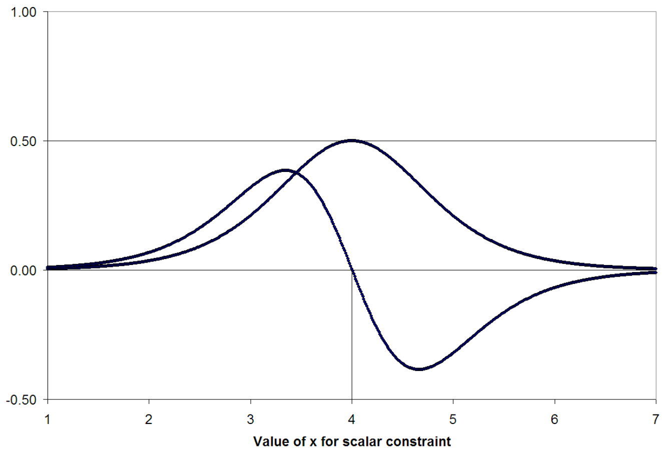

Figure 4

- The second derivative has an intriguing symmetrical pattern: just a little bit above zero coming up from negative infinity, peaking, then plunging rapidly to its lowest value, and lastly edging upward to remain just a little bit below zero out to positive infinity. The point at which the plunge reaches zero looks to be at $x = 4$; let us confirm.

- The second derivative of our logistic function:

  $$\frac{1}{1 + e^{F - Mx}}$$

  turns out to be:

  $$\frac{d^2y}{dx^2} = \frac{M^2 e^{F - Mx}(e^{F - Mx} - 1)}{(1 + e^{F - Mx})^3}$$

  Again, see §9 for how this can be calculated.

- The last step is to plug the symmetry point, $x = F/M$, into our formula for the second derivative. If F/M is a maximum (or minimum, really, but as you can see it turned out to be a maximum) this ought to yield zero.

  $$\begin{aligned}
  \frac{d^2y}{dx^2}(F/M) &= \frac{M^2 e^{F - M \times F/M}(e^{F - M \times F/M} - 1)}{(1 + e^{F - M \times F/M})^3} \\
  &= \frac{M^2 e^{F - F}(e^{F - F} - 1)}{(1 + e^{F - F})^3} \\
  &= \frac{M^2 e^{0}(e^{0} - 1)}{(1 + e^{0})^3} \\
  &= \frac{M^2 1(1 - 1)}{(1 + 1)^3} \\
  &= \frac{M^2 1(0)}{(2)^3} \\
  &= 0
  \end{aligned}$$

  Sure enough.

- Summarising: the second derivative of the logistic function

  $$y = \frac{1}{1 + e^{F - Mx}} \quad \text{is} \quad \frac{M^2 e^{F - Mx}(e^{F - Mx} - 1)}{(1 + e^{F - Mx})^3}$$

  This reaches zero at M/F, the symmetry point of the logistic function. This means that the first derivative, representing the slope of the logistic function, reaches an extreme at M/F, which by inspection we see is the maximum. The maximum slope occurs at M/F.

9 Finding the derivatives

9.1 Reviewing the rules for differentiation in calculus

9.1.1 *Derivative of a constant*

If $y(x) = \mathrm{c}$, where c is some constant, then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = 0$$

9.1.2 *Derivative of a linear equation*

If $y(x) = \mathrm{m}x + \mathrm{b}$, where m and b are constants, then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \mathrm{m}$$

9.1.3 *Differentiating a power*

If $y = x^{\mathrm{n}}$, then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \mathrm{n}x^{\mathrm{n}-1}$$

It's fine for n to be negative, since $1/x^{\mathrm{n}}$ is the same as $x^{-\mathrm{n}}$. We will see this below.

9.1.4 *Differentiating exponentials*

If $y = \mathrm{e}^{x}$, then, amazingly,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \mathrm{e}^{x}$$

9.1.5 *Sum Rule*

The derivative of the sum of two functions is the sum of their derivatives.

9.1.6 *Product Rule*

If $y = \mathrm{f}(x) \times \mathrm{g}(x)$, then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = (\mathrm{f}(x) \times \frac{\mathrm{dg}}{\mathrm{d}x}) + (\mathrm{g}(x) \times \frac{\mathrm{df}}{\mathrm{d}x})$$

9.1.7 *Chain Rule*

If $y = \mathrm{f}(\mathrm{g}(x))$, then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{df}}{\mathrm{dg}} \times \frac{\mathrm{dg}}{\mathrm{d}x}$$

This can be applied more than once, so if $y = \mathrm{f}(\mathrm{g}(\mathrm{h}(x)))$, then

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{df}}{\mathrm{dg}} \times \frac{\mathrm{dg}}{\mathrm{dh}} \times \frac{\mathrm{dh}}{\mathrm{d}x}$$

9.2 Finding the first derivative of the logistic function

9.2.1 *Setting up the problem for solution with the Chain Rule*

- We seek to differentiate our logistic function, which is:

$$y = \frac{1}{1 + \mathrm{e}^{\mathrm{F} - \mathrm{M}x}}$$

- This is facilitated if we rewrite it as a 'function of a function of a function', with intermediate steps.

$$y = \mathrm{f}(\mathrm{g}) = \frac{1}{\mathrm{g}}$$

$$\mathrm{g}(\mathrm{h}) = 1 + \mathrm{e}^{\mathrm{h}}$$

$$\mathrm{h}(x) = \mathrm{F} - \mathrm{M}x$$

- In other words, we work through the formula and re-express its content step-by-step. The functions $\mathrm{h}(x)$, $\mathrm{g}(\mathrm{h})$ and $\mathrm{f}(\mathrm{g})$ re-express the formula going 'from the inside out'. We can put the functions back together as follows:

$$y = \mathrm{f}(\mathrm{g}(\mathrm{h}(x)))$$

- The Chain Rule (9.1.7) tells us to take the derivative of each 'sub-function', then multiply the derivatives out to get the whole derivative.

9.2.2 *Differentiating* f(g), g(h) *and* h(x)

- If

$$\mathrm{f}(\mathrm{g}) = \frac{1}{\mathrm{g}}$$

then

$$\frac{\mathrm{df}}{\mathrm{dg}} = \frac{-1}{\mathrm{g}^2}$$

This applies (9.1.3), treating $1/\mathrm{g}$ as its equivalent g^{-1}.

- If $\mathrm{g}(\mathrm{h}) = 1 + \mathrm{e}^{\mathrm{h}}$, then we have a sum of two functions; a trivial one (the constant 1) and e^{h}.
  - We differentiate them separately, then add (9.1.5).
  - The derivative of the constant one is zero (9.1.1).
  - The derivative of e^{h} is itself (9.1.4).

- so:

$$\frac{dg}{dh} = e^h$$

- If $h(x) = F - Mx$ then by 9.1.2 we have

$$\frac{dh}{dx} = -M$$

9.2.3 *Applying the Chain Rule*

- Now that we have our three derivatives, we can substitute them back into the formula of the Chain Rule:

$$\frac{dy}{dx} = \frac{df}{dg} \times \frac{dg}{dh} \times \frac{dh}{dx}$$

$$= \frac{-1}{g^2} \times e^h \times -M$$

- And we know what g and h are (9.2.1, second bullet point), so we can substitute their definitions back in to the formula:

$$\frac{dy}{dx} = \frac{-1}{(1 + e^h)^2} \times e^{F - Mx} \times -M$$

And once again, filling in the definition of $h(x)$:

$$= \frac{-1}{(1 + e^{F - Mx})^2} \times e^{F - Mx} \times -M$$

- Clean-up: the two minus signs cancel each other, and we re-arrange prettily into a single fraction:

$$\frac{dy}{dx} = \frac{Me^{F - Mx}}{(1 + e^{F - Mx})^2}$$

- This is the correct answer, and is indeed the derivative assumed in 7.1, second bullet point.

9.2.4 *Checking the answer*

- Are we sure we're right? A check comes from the process of making the actual graphs displayed above. These are made (in Excel) by plotting one dot at every interval of 0.01 along the x axis. The fraction

$$\frac{y(x_{n+1}) - y(x_n)}{x_{n+1} - x_n}$$

where x_n is the nth value along the x axis, is a close approximation to the derivative when the interval between dots is small. We find that the values thus obtained closely match those calculated with the true derivative

$$\frac{\mathrm{M}e^{\mathrm{F}-\mathrm{M}x}}{(1+e^{\mathrm{F}-\mathrm{M}x})^2}$$

thus confirming that we haven't made a mistake.

9.3 Finding the second derivative of the logistic function

9.3.1 *Defining the task*

- To obtain the second derivative, we need to differentiate the first derivative we just obtained, i.e. we want dy/dx, where $y(x)$ is now redefined as

  $$y(x) = \frac{\mathrm{M}e^{\mathrm{F}-\mathrm{M}x}}{(1+e^{\mathrm{F}-\mathrm{M}x})^2}$$

- This derivative has two appearances of x in it, so we can't follow the same method we used for the first derivative. Rather, our strategy is to treat

  $$\frac{\mathrm{M}e^{\mathrm{F}-\mathrm{M}x}}{(1+e^{\mathrm{F}-\mathrm{M}x})^2}$$

  as a product, and use the Product Rule (9.1.6).

9.3.2 *Setting up for the Product Rule*

The product is this:

$$y(x) = \mathrm{M}e^{\mathrm{F}-\mathrm{M}x} \times \frac{1}{(1+e^{\mathrm{F}-\mathrm{M}x})^2}$$

- We'll give the multiplicands names, so we can refer to them.

  $$f(x) = \mathrm{M}e^{\mathrm{F}-\mathrm{M}x}$$

  $$g(x) = \frac{1}{(1+e^{\mathrm{F}-\mathrm{M}x})^2}$$

- The Product Rule tells us that we will get our derivative if we compute the derivatives of $f(x)$ and $g(x)$ and plug them into the formula

  $$\frac{dy}{dx} = (f(x) \times \frac{dg}{dx}) + (g(x) \times \frac{df}{dx})$$

- This requires that we differentiate both $f(x)$ and $g(x)$.

9.3.3 *Differentiating* $f(x)$

- We have

  $$f(x) = \mathrm{M}e^{\mathrm{F}-\mathrm{M}x}$$

- This can be set up for application of the Chain Rule again, with functions h(k), k(m), m(x).

$h(k) = Mk$

$k(m) = e^m$

$m(x) = F - Mx$

- The derivatives are:

$\frac{dh}{dk} = M$ (by 9.1.2, where the b in 9.1.2 is taken to be zero)

$\frac{dk}{dm} = e^m$ (by 9.1.4)

$\frac{dm}{dx} = -M$ (by 9.1.2)

- Multiplying it out, following the Chain Rule:

$$\frac{df}{dx} = \frac{dh}{dk} \times \frac{dk}{dm} \times \frac{dm}{dx}$$

$$= M \times e^m \times -M$$

$$= -M^2 e^m$$

- Substituting in the definition of m(x):

$$= -M^2 e^{F - Mx}$$

This is the derivative of f(x).

9.3.4 *Differentiating* g(x)

- We have:

$$g(x) = \frac{1}{(1 + e^{F-Mx})^2}$$

- This can be set up for application of the Chain Rule again, with functions h(k), k(m), m(x) (we are recycling some letters here).

$h(k) = \frac{1}{k^2}$

$k(m) = 1 + e^m$

$m(x) = F - Mx$

- The derivatives:

$\frac{dh}{dk} = \frac{-2}{k^3}$ (by 9.1.3, where the power in question is −2)

$\frac{dk}{dm} = e^m$ (sum of the derivatives of the constant 1 and of e^m; see 9.1.5, 9.1.1 and 9.1.4)

$\frac{dm}{dx} = -M$ (by 9.1.2)

- Multiplying it out, following the Chain Rule:

$$\frac{df}{dx} = \frac{dh}{dk} \times \frac{dk}{dm} \times \frac{dm}{dx}$$

$$= \frac{-2}{k^3} \times e^m \times -M$$

$$= \frac{2Me^m}{k^3}$$

- Substituting in the definition of $k(x)$:

$$= \frac{2Me^m}{(1 + e^m)^3}$$

- Substituting in the definition of $m(x)$:

$$\frac{dg}{dx} = \frac{2Me^{F-Mx}}{(1 + e^{F-Mx})^3}$$

This is the derivative of $g(x)$.

9.3.5 *Applying the Product Rule*

- We now have the functions and derivatives we need to plug into the Product Rule formula:

$$\frac{dy}{dx} = (f(x) \times \frac{dg}{dx}) + (g(x) \times \frac{df}{dx})$$

- All four expressions we need appear above. Hunting them down and plugging them in, we get:

$$\frac{dy}{dx} = \left(Me^{F-Mx} \times \frac{2Me^{F-Mx}}{(1 + e^{F-Mx})^3}\right) + \left(\frac{1}{(1 + e^{F-Mx})^2} \times -M^2e^{F-Mx}\right)$$

This is indeed the derivative, but it looks like a mess.

- Let us clean it up a bit. Factor out M^2 (2.1c):

$$\frac{dy}{dx} = M^2 \times \left((e^{F-Mx} \times \frac{2e^{F-Mx}}{(1 + e^{F-Mx})^3}) + (\frac{1}{(1 + e^{F-Mx})^2} \times -e^{F-Mx})\right)$$

- Factor out $1/(1 + e^{F-Mx})^2$:

$$\frac{dy}{dx} = M^2 \times \frac{1}{(1 + e^{F-Mx})^2} \times \left((e^{F-Mx} \times \frac{2e^{F-Mx}}{1 + e^{F-Mx}}) + -e^{F-Mx}\right)$$

- Make a pretty fraction:

$$\frac{dy}{dx} = \frac{M^2}{(1 + e^{F - Mx})^2} \times \left((e^{F - Mx} \times \frac{2e^{F - Mx}}{1 + e^{F - Mx}}) + -e^{F - Mx}\right)$$

- Factor out $e^{F - Mx}$:

$$\frac{dy}{dx} = \frac{M^2}{(1 + e^{F - Mx})^2} \times e^{F - Mx} \times \left(\frac{2e^{F - Mx}}{1 + e^{F - Mx}} - 1\right)$$

- Make a pretty fraction:

$$\frac{dy}{dx} = \frac{M^2 e^{F - Mx}}{(1 + e^{F - Mx})^2} \times \left(\frac{2e^{F - Mx}}{1 + e^{F - Mx}} - 1\right)$$

- Replace the 1 by the equivalent $(1 + e^{F - Mx})/(1 + e^{F - Mx})$:

$$\frac{dy}{dx} = \frac{M^2 e^{F - Mx}}{(1 + e^{F - Mx})^2} \times \left(\frac{2e^{F - Mx}}{1 + e^{F - Mx}} - \frac{1 + e^{F - Mx}}{1 + e^{F - Mx}}\right)$$

- Gather the terms with the same denominator:

$$\frac{dy}{dx} = \frac{M^2 e^{F - Mx}}{(1 + e^{F - Mx})^2} \times \left(\frac{2e^{F - Mx} - (1 + e^{F - Mx})}{1 + e^{F - Mx}}\right)$$

- Subtract:

$$\frac{dy}{dx} = \frac{M^2 e^{F - Mx}}{(1 + e^{F - Mx})^2} \times \frac{e^{F - Mx} - 1}{1 + e^{F - Mx}}$$

- Multiply out the whole thing so we get a single fraction:

$$\frac{dy}{dx} = \frac{M^2 e^{F - Mx}(e^{F - Mx} - 1)}{(1 + e^{F - Mx})^3}$$

- This is about as clean as we can make it. This is the second derivative of our logistic function, as employed in 8.2 above to verify the location of maximum slope.

- It checks by the same method we used for the first derivative.

Appendix B: Three hypothetical curves fittable using domain-indexed constraints, but ill-fitted by the scalar model

These graphs are meant to demonstrate the excess descriptive power of the domain-indexation theory discussed in §7.2 of the paper. In all three graphs, the black line labelled 'observed' represents a hypothetical data pattern that, like all monotonic patterns, can be derived under the domain-indexation theory. The grey line is the best-fit curve under our scalar model. It can be seen that in each case the scalar model performs poorly – as we believe it should – in fitting the hypothetical data.

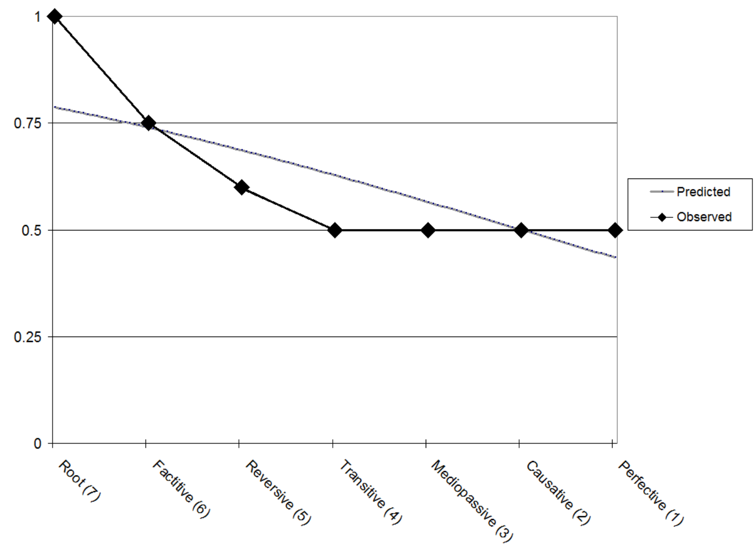

Figure 5

Starts at one, asymptotes at 0.5.

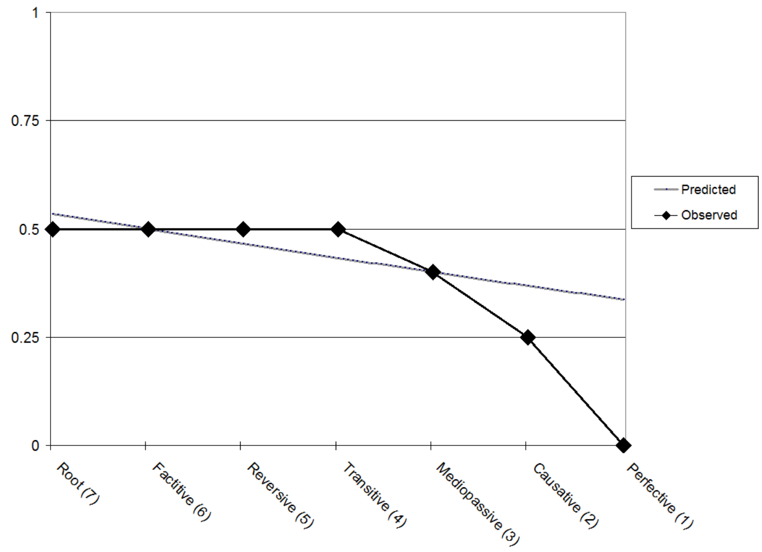

Figure 6
Starts with asymptote at 0.5, descends to zero.

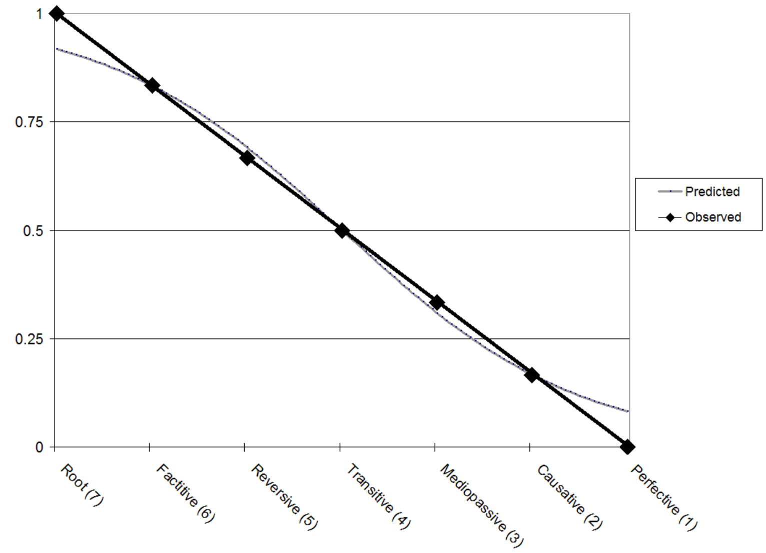

Figure 7
Linear descent from one to zero.

Appendix C: Testing type *vs*. token variation in Tommo So vowel harmony

1.1 Predictions of theories

- Pure type-variation theory
    - Every stem + suffix combination has an invariant outcome, reflecting lexical listing (or some mechanism, like diacritics, with comparable function).
- Pure token-variation theory
    - Every stem + suffix combination, if we were able to get speakers to say it many times, would vary, harmonising at a rate corresponding to the base rate for harmony with the particular vowels and suffix it contains.
- Mixed theories
    - There is *some* lexical listing, and some online generation.

1.2 The data that could bear on the token *vs*. type issue

- We need stem + suffix combinations that:
    - appear more than once in the data (otherwise variation is not detectible)
    - involve a particular instance of harmony with a probability greater than zero and less than one (otherwise variation not possible).

1.3 The data

- We found 155 stem + suffix combinations that fit this description.
- They form 723 tokens in total.
- For each, we have three numbers:
    - Total tokens
    - Of these, how many harmonise
    - The theoretical rate of vowel harmony, estimated from the full data set

token count	harmo -nised	*p*(harmony)	token count	harmo -nised	*p*(harmony)	token count	harmo -nised	*p*(harmony)	token count	harmo -nised	*p*(harmony)
48	48	0.990	5	0	0.440	3	0	0.440	2	0	0.440
30	10	0.136	5	0	0.214	3	0	0.440	2	0	0.440
24	20	0.851	5	0	0.197	3	0	0.214	2	0	0.440
24	11	0.690	4	4	0.909	3	0	0.181	2	0	0.440
17	8	0.440	4	4	0.690	3	0	0.181	2	0	0.440
16	2	0.136	4	4	0.440	3	0	0.136	2	0	0.440
16	0	0.214	4	4	0.440	2	2	0.990	2	0	0.440
15	2	0.197	4	4	0.197	2	2	0.990	2	0	0.440
13	13	0.440	4	3	0.990	2	2	0.990	2	0	0.440
12	10	0.851	4	3	0.909	2	2	0.990	2	0	0.440
11	7	0.851	4	3	0.690	2	2	0.909	2	0	0.440
11	4	0.440	4	3	0.440	2	2	0.909	2	0	0.440
11	0	0.214	4	2	0.440	2	2	0.909	2	0	0.440
10	10	0.851	4	0	0.440	2	2	0.909	2	0	0.440
10	4	0.214	4	0	0.440	2	2	0.851	2	0	0.440
10	3	0.136	4	0	0.214	2	2	0.851	2	0	0.214
10	0	0.214	4	0	0.181	2	2	0.690	2	0	0.214
8	8	0.990	4	0	0.136	2	2	0.690	2	0	0.214
8	0	0.136	4	0	0.136	2	2	0.440	2	0	0.197
7	7	0.990	3	3	0.990	2	2	0.440	2	0	0.197
7	7	0.440	3	3	0.990	2	2	0.440	2	0	0.197
7	6	0.440	3	3	0.990	2	2	0.214	2	0	0.197
7	3	0.136	3	3	0.990	2	2	0.214	2	0	0.197
7	2	0.136	3	3	0.909	2	2	0.181	2	0	0.197
7	1	0.136	3	3	0.851	2	1	0.690	2	0	0.197
6	6	0.909	3	3	0.690	2	1	0.690	2	0	0.181
6	6	0.440	3	3	0.690	2	1	0.690	2	0	0.181
6	5	0.214	3	3	0.440	2	1	0.440	2	0	0.181
6	3	0.440	3	3	0.440	2	1	0.440	2	0	0.181
6	1	0.214	3	3	0.440	2	1	0.440	2	0	0.181
6	0	0.214	3	3	0.197	2	1	0.440	2	0	0.136
6	0	0.197	3	3	0.185	2	1	0.440	2	0	0.136
6	0	0.062	3	3	0.185	2	1	0.440	2	0	0.136
5	5	0.990	3	2	0.181	2	1	0.440	2	0	0.136
5	5	0.440	3	1	0.440	2	1	0.197	2	0	0.136
5	5	0.214	3	1	0.440	2	1	0.181	2	0	0.136
5	4	0.440	3	1	0.181	2	0	0.440	2	0	0.136
5	1	0.440	3	0	0.440	2	0	0.440	2	0	0.136
5	0	0.440	3	0	0.440	2	0	0.440			

2 Test I: Monte Carlo simulation

2.1 Procedure

- Take all the 723 data in 1.3.
- For each datum, use a random number generator to perform a random trial, deriving a 'harmony' or 'non-harmony' outcome according to the probability of harmony for this suffix/stem vowel combination.
- Count how many 'always harmonising' words show up.
- Count how many 'never harmonising' words show up
- Repeat above 100,000 times to obtain an estimated probability distribution.
- Then see where the real values fit into this distribution. This yields a *p*-value.
- We programmed this procedure in software (code below).

2.2 Result

number of cases of no-harmony	number in Monte Carlo simulation	real value	number of cases of all-harmony	number in Monte Carlo simulation	real value
16	0		16	0	
17	0		17	0	
18	0		18	0	
19	0		19	6	
20	0		20	17	
21	0		21	53	
22	0		22	136	
23	1		23	346	
24	4		24	807	
25	7		25	1565	
26	8		26	2799	
27	17		27	4541	
28	44		28	6649	
29	99		29	8764	
30	171		30	10571	
31	288		31	11531	
32	544		32	11698	
33	897		33	10756	
34	1339		34	8951	
35	2097		35	7145	
36	3092		36	5216	
37	3860		37	3496	

38	5085		38	2308	
39	6367		39	1268	
40	7329		40	708	
41	8022		41	377	
42	8407		42	161	
43	8532		43	83	
44	8333		44	36	
45	7616		45	6	
46	6737		46	3	
47	5693		47	3	
48	4555		48	0	actual value: 48
49	3457		49	0	
50	2538		50	0	
51	1759		51	0	
52	1169		52	0	
53	788		53	0	
54	514		54	0	
55	295		55	0	
56	147		56	0	
57	100		57	0	
58	61		58	0	
59	16		59	0	
60	7		60	0	
61	3		61	0	
62	2		62	0	
63	0		63	0	
64	0		64	0	
65	0		65	0	
66	0	actual value: 66	66	0	
67	0		67	0	
68	0		68	0	
69	0		69	0	
70	0		70	0	
71 +	0		71 +	0	

- For both all-harmony and no-harmony, the actual value falls beyond all outcomes obtained in 100,000 Monte Carlo trials. Therefore, $p < 0.00001$.
- This looks like a very powerful type effect.

2.3 An alternative hypothesis to explain the effect

- Many of the data are just *two tokens* – perhaps consultants 'self-prime' when repeating a word in short time interval?
- This would in itself produce more unanimous 'all harmony' or 'no harmony' forms than we would expect by chance.

3 Test II: binomial test

3.1 Rationale

- Suppose there really are individual stem + suffix combinations that are lexically listed (perhaps with some particular level of 'strength') to come out all-harmony or no-harmony.
- Then we should be able to detect these cases: they will typically have far more (or far fewer) cases of harmony than we would expect from the baseline probability of harmony in their category (i.e. their vowels and suffix).
- Such disparities can be submitted to significance testing.

3.2 Testing for candidate listed items

- We have n tokens, of which m are harmonised, in a harmony condition with probability p.
- The odds of getting m or fewer are given by the cumulative binomial distribution, which in Excel is:

  = BINOMDIST(m, n, p, true), where 'true' means 'cumulative'

- And the probability of getting m or more is

  = 1 − BINOMDIST($m - 1$, n, p, true)

3.3 Top ten forms for ‘surprisingly infrequent harmony’

Form	*p*(*harmony*)	*n*	*m*	*p*(*observing this few*)
ɔ̀g-TRANS	0.690	24	11	0.015
gòó-PERF	0.214	16	0	0.021
jɔ̀bɔ́-FACT	0.990	4	3	0.039
nɔ́gɔ́mɔ́-MEDIOPASS	0.440	5	0	0.055
mùnnɔ́-MEDIOPASS	0.440	5	0	0.055
jáá-FACT	0.851	11	7	0.068
yóó-PERF	0.214	11	0	0.071
sɔ́ɔ́-PERF	0.214	10	0	0.090
mɔ̀ɔ̀mbɔ́-MEDIOPASS	0.440	4	0	0.098
ùsú-FACT.MEDIOPASS	0.440	4	0	0.098

- This seems very unimpressive, given that we didn’t correct for ‘fishing’.

3.4 Top ten forms for ‘surprisingly frequent harmony’

Form	*p*(*harmony*)	*n*	*m*	*p*(*observing this many*)
ɲ̀jó-MEDIOPASS	0.440	13	13	2.32e-05
párá-REVERS	0.197	4	4	0.002
súgó-PERF	0.214	6	5	0.002
dùyɔ́-MEDIOPASS	0.440	7	7	0.003
óbó-PERF	0.136	30	10	0.005
tɛ́mɛ́-CAUS	0.185	3	3	0.006
yɛ̀-CAUS	0.185	3	3	0.006
dùló-MEDIOPASS	0.440	6	6	0.007
nárá-REVERS	0.197	3	3	0.008
wòró-MEDIOPASS.CAUS	0.440	5	5	0.016

- This looks a little more impressive.

3.5 Our surmise: the 'treat as if one word' theory

- Often in phonology a word of multiple parts gets pronounced as if one: *cupboard* ≠ *cup* + *board*
- More subtly, English *high school* is durationally one word; *pie school* is durationally two.
- We suspect that the words in § 3.4 are 'treated as one word' — this explains their harmony, because simple unsuffixed polysyllabic words in Tommo So always obey harmony.

3.6 Upshot

- We think the most likely explanation of what is going on would follow three hypotheses:
  - ➢ As a first approximation, Tommo So vowel harmony involves token variation; the dice are rolled anew at each utterance. This explains why we find so few cases of words whose harmony distribution deviates significantly from random.
  - ➢ There is, however, a self-priming effect, which led to greater agreement across elicitations for any given word than would be expected by chance (Monte Carlo simulation).
  - ➢ Lastly, there is a small number of words that are lexically listed as single entries, essentially as stems, as in English *high school*; as such, these emerge with unanimous harmony, testing as significant in Test 2.

4 Code for the Monte Carlo simulation

4.1 Details

Written in Visual Basic 6. The code is probably generic enough to be read by any programmer. Windows executable available from the authors on request.

4.2 Code

```
'Monte Carlo.

'Run the Monte Carlo simulation reported in McPherson & Hayes: "Relating application frequency to
morphological structure: the case of Tommo So vowel harmony". Phonology 33.

Option Explicit

Private Sub StartButton_Click()

  'This routine runs when the user clicks the Start button on the interface.

  'Variable declarations

      'Empirical variables about the input file.
          Dim NumberOfTokens(155)       'Number of tokens attested for each of 155 inputs.
          Dim NumberOfHits(155)         'Number of hits (harmony)
          Dim Probability(155)          'The probability of harmony for a word with these vowels and suffix,
                                        '  according to the McPherson/Hayes analysis.
          Dim NumberOfRealAllHarmonyCases As Long  'These get counted by the program.
          Dim NumberOfRealNoHarmonyCases As Long

      'Variables for file input.
          Dim FileHeader As String
          Dim MyLine As String
          Dim NumberOfEntries As Long

      'Variables for the Monte Carlo simulation
          'Count how many harmony outcomes you get for one single input.
```

```
            Dim LocalHitCount As Long
        'Count how many all-harmony cases and no-harmony cases you get in one single simulation of the
                    data.
            Dim NumberAllHarmony As Long
            Dim NumberNoHarmony As Long
        'Count how many simulations result in each possible total
            Dim OutputDistributionAllHarmony(155) As Long
            Dim OutputDistributionNoHarmony(155) As Long

    'Indices used in various loops.
            Dim EntryIndex As Long, TrialIndex As Long, TokenIndex As Long

'Open the input and output files.
    Open App.Path + "/S0952675716000051sup008.txt" For Input As #1
    Open App.Path + "/S0952675716000051sup009.txt" For Output As #2

'Read the input file.
    'Skip the header line
        Line Input #1, MyLine
    'The data:
        Do While Not EOF(1)  'EOF is 'end of file'
            Line Input #1, MyLine
            'Ignore an index number using the residue (post-tab) function.
                Let MyLine = s.Residue(MyLine)
            'Record what entry you are working on:
                Let NumberOfEntries = NumberOfEntries + 1
            'Grab the number of tokens for this entry, using the (pre-tab) chomp function.
                Let NumberOfTokens(NumberOfEntries) = Val(s.Chomp(MyLine))
            'Grab the number of hits.
                Let MyLine = s.Residue(MyLine)
                Let NumberOfHits(NumberOfEntries) = Val(s.Chomp(MyLine))
            'The rest of the line is the probability.
                Let Probability(NumberOfEntries) = Val(s.Residue(MyLine))
        Loop

'Calculate for real life the number of no-harmony and all-harmony outcomes.
```

```
    For EntryIndex = 1 To NumberOfEntries
        If NumberOfHits(EntryIndex) = NumberOfTokens(EntryIndex) Then
            Let NumberOfRealAllHarmonyCases = NumberOfRealAllHarmonyCases + 1
        ElseIf NumberOfHits(EntryIndex) = 0 Then
            Let NumberOfRealNoHarmonyCases = NumberOfRealNoHarmonyCases + 1
        End If
    Next EntryIndex

'Do 100,000 simulations.
    For TrialIndex = 1 To 100000
        'Report progress so user will know the program is running.
            If TrialIndex Mod 10000 = 0 Then
                Let StartButton.Caption = "Completed " + Str(TrialIndex) + " trials"
            End If
        'Initialize the variables the record the number of all-harmony and no-harmony outcomes.
            Let NumberAllHarmony = 0
            Let NumberNoHarmony = 0
        'Go through the data
            For EntryIndex = 1 To NumberOfEntries
                'Initialize the variable recording how much harmony you got.
                    Let LocalHitCount = 0
                'Simulate n utterances, n the number of tokens for this entry.
                    For TokenIndex = 1 To NumberOfTokens(EntryIndex)
                        'A hit is when the 0-1 random variable comes out less than the probability.
                            If Rnd() < Probability(EntryIndex) Then
                                'Augment number of hits.
                                    Let LocalHitCount = LocalHitCount + 1
                            End If
                    Next TokenIndex
                'Determine if this was unanimous (for, or against, harmony), and augment the counts for
                        unanimous and zero cases.
                    If LocalHitCount = 0 Then
                        Let NumberNoHarmony = NumberNoHarmony + 1
                    ElseIf LocalHitCount = NumberOfTokens(EntryIndex) Then
                        Let NumberAllHarmony = NumberAllHarmony + 1
                    End If
```

```
            Next EntryIndex
        'This completes a trial.  Record the result by augmenting the entries for number of all-harmony
             and no-harmony.
            Let OutputDistributionAllHarmony(NumberAllHarmony) =
                    OutputDistributionAllHarmony(NumberAllHarmony) + 1
            Let OutputDistributionNoHarmony(NumberNoHarmony) =
                    OutputDistributionNoHarmony(NumberNoHarmony) + 1
    Next TrialIndex

'Print what you learned.
    'Header:
        Print #2, "Number"; Chr(9); "No harmony"; Chr(9); "All harmony"
    'Go through cases:
        For EntryIndex = 0 To 155
            Print #2, EntryIndex;
            'Note where the real count lies:
                Print #2, Chr(9); OutputDistributionNoHarmony(EntryIndex);
                If EntryIndex = NumberOfRealNoHarmonyCases Then
                    Print #2, Chr(9); "(real)";
                Else
                    Print #2, Chr(9);
                End If
            Print #2, Chr(9); OutputDistributionAllHarmony(EntryIndex);
            'Note where the real count lies:
                If EntryIndex = NumberOfRealAllHarmonyCases Then
                    Print #2, Chr(9); "(real)";
                End If
            'End of line:
                Print #2,
        Next EntryIndex

'Finish up
    Close
    End
```

• `End Sub`