An identity preference in Ngbaka vowels

Juliet Stanton, NYU (stanton@nyu.edu)

AMP 2021
October 2, 2021

I. Overview

- Models of long-distance phonology differ in whether they propose an explicit proviso for identity.
  - Gallagher & Coon (2009): IDENTITY, which requires consonants linked by a correspondence relation to be identical.
  - Rose & Walker (2004) and other work in ABC: identity not a goal; no explicit proviso for identity.
  - Zuraw (2002): identity is a goal; no explicit proviso for identity.
- **A question** that this raises: should identity preferences be analyzed using one monolithic constraint, or a set of constraints, each requiring identity for an individual feature?
- **This poster:** a case study of Ngbaka vowels, which suggests that we need both.

II. Background and research question

- Ngbaka (Ubangian; Thomas 1963, Henrix et al. 2007, Sélézilo 2008, *a.o.*) has a twelve-vowel system.

Oral vowels

i		u
e		o
ɛ		ɔ
	ɑ	

Nasal vowels

ĩ		ũ
ɛ̃		ɔ̃
	ɑ̃	

- Descriptions of Ngbaka (references above) often claim that multiple types of vowel harmony are active. Two examples:
  - ATR harmony: [e ẽ o õ] cannot precede or follow [ɛ ɛ̃ ɔ ɔ̃].
  - Backness harmony: [i ĩ e ɛ ɛ̃] cannot precede or follow [u ũ o ɔ ɔ̃].
- These requirements often coincide to ensure that, in a CV_1CV_2 word, V_1 and V_2 are identical (noted by Sélézilo 2008). Another analytical possibility arises: Ngbaka vowels *prefer to be identical.*
- ***Question:*** can the apparent preference for identity in Ngbaka vowels be explained through the interaction of multiple types of harmony, or must an identity preference be recognized independently?

III. Data

- To address this question, I created a database of vowel pairs from Henrix et al.'s (2015) Ngbaka dictionary (5,571 words).
- Investigation limited to disyllabic or longer words that contain only CV syllables (3,928 words).
- For each word, I extracted each adjacent pair of vowels.

z **i** b **$ɔ_1$** l **$ɔ_2$**
Pairs: [**i $ɔ_1$**], [**$ɔ_1$ $ɔ_2$**]

t **ɑ** l **ɛ** k **u** s **i**
Pairs: [**ɑ ɛ**], [**ɛ u**], [**u i**]

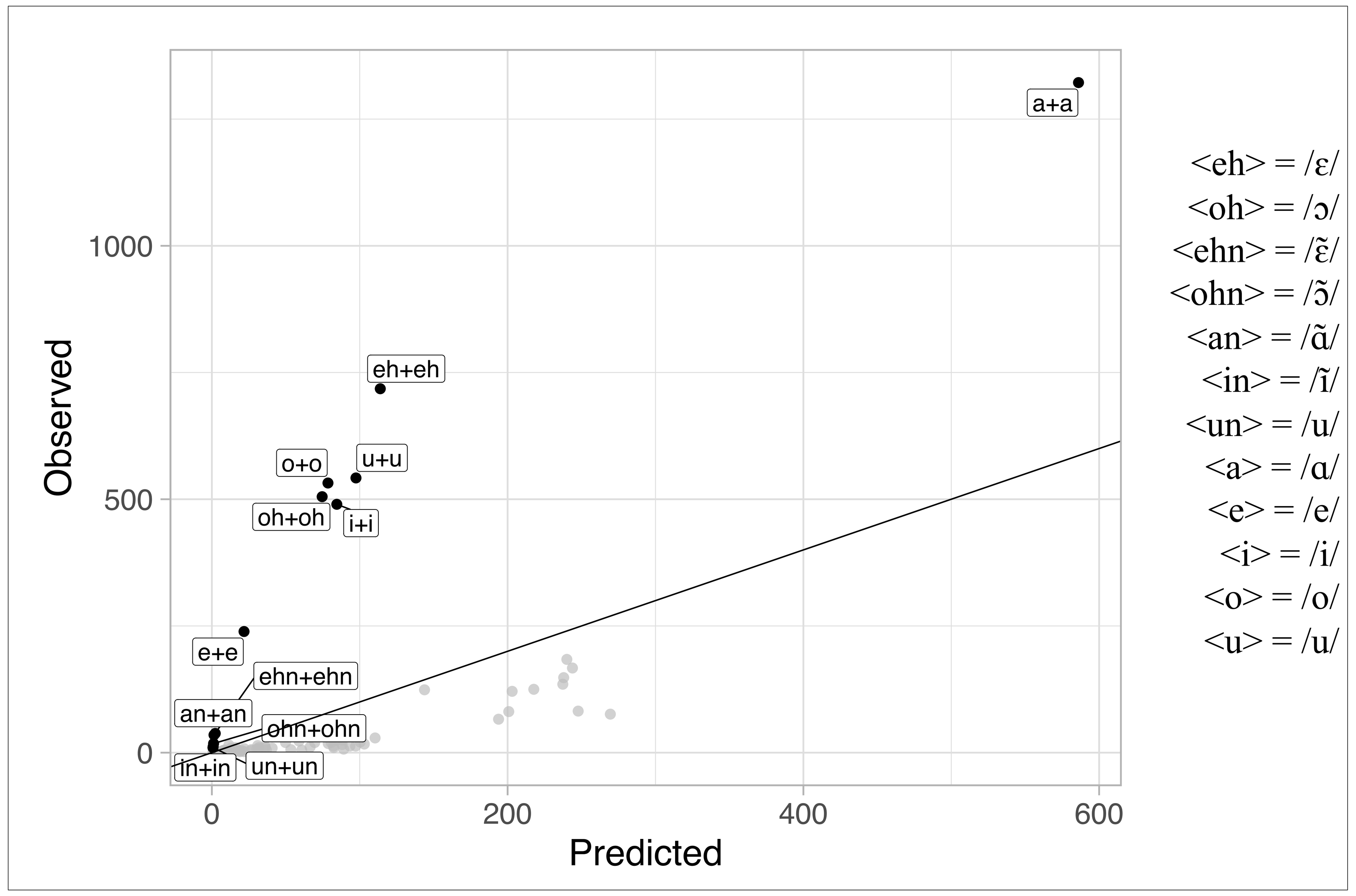
- **Result**: 6,716 vowel pairs. Identical pairs (n=4,461, in white) are more common than non-identical pairs (n=2,255, in black).

		V_2											
		ɑ	**ɑ̃**	**e**	**ɛ**	**ɛ̃**	**i**	**ĩ**	**o**	**ɔ**	**ɔ̃**	**u**	**ũ**
V_1	**ɑ**	1322	33	124	76	10	167	6	135	81	14	148	8
	ɑ̃	7	38	1	0	0	9	9	4	3	1	3	1
	e	26	0	239	5	1	5	0	13	7	1	10	0
	ɛ	82	3	9	718	6	17	0	21	24	2	33	0
	ɛ̃	4	4	0	1	35	1	1	2	1	1	3	4
	i	121	6	20	13	2	490	1	59	20	4	10	1
	ĩ	2	2	2	0	0	1	12	0	3	4	0	0
	o	66	2	55	7	1	20	0	532	11	1	18	0
	ɔ	125	10	6	47	3	44	5	16	505	15	25	0
	ɔ̃	2	2	0	4	0	2	5	0	4	18	1	0
	u	184	5	23	29	4	62	3	13	13	1	542	4
	ũ	4	2	0	0	5	1	4	1	0	0	5	10

- An interesting trend to note: if a vowel co-occurs with a non-identical vowel, it's usually [ɑ].
- In addition, there are trends in these data that I won't focus on (example: nasal vowels are less common than oral vowels).

IV. Analysis

- Analysis: loglinear model of count data, using *bayesglm* from R's arm package (Gelman & Hill 2007).
- Baseline model asks: **what is the predicted frequency of each vowel pair, given the independent frequency of each vowel?**
  - Dependent variable: number of times a particular vowel-vowel pair is attested.
  - Independent variables: one predictor per vowel per position ($ɑ_1$, $ɑ_2$, $ɑ̃_1$, $ɑ̃_2$, e_1, e_2, etc.).
- Fitted values (obtained with R's *fitted.values* function) show that identical pairs are overattested relative to expectation.



For an interactive graph where all points are labeled, go to:

julietstanton.github.io/files/ngbakavowels.html

- To this baseline model, I added predictors reflecting types of vowel harmony and one predictor to look for effects of identity.

	Predictor (all binary)	Type of harmony	Assigns a 1 to....
a.	*[αATR][-αATR]	ATR	(e|o)(ɛ|ɔ) and (ɛ|ɔ)(e|o)
b.	*[αback][-αback]	Backness	(i|e|ɛ)(u|o|ɔ) and (u|o|ɔ)(i|e|ɛ)
c.	*[αhigh, -back, -low][-αhigh, -back, -low]	Height amongst front vowels	(i)(e|ɛ) and (e|ɛ)(i)
d.	*[αhigh, +back, -low][-αhigh, +back, -low]	Height amongst back vowels	(u)(o|ɔ) and (o|ɔ)(u)
e.	*[αnasal][-αnasal]	Nasal	vowels mismatching for nasality
f.	$*[+syllabic]_i[+syllabic]_j$	Identity	non-identical vowel pairs

- ***Question:*** which combination of predictors is responsible for shaping the data?
- To answer this question, I fit the maximal model to the data and compared the goodness of fit of nested models using LRTs.

V. Results

- The best-fit model includes predictors for ATR harmony (a; t = 5.51, $p < .001$), backness harmony (b; t = 4.38, $p < .001$); height harmony among front vowels (c; t = 5.07, $p < .001$), nasal harmony (e; t = 6.64, $p < .001$), and identity (f; t = 29.23, $p < .001$).
  - Further modeling suggests another possible type of harmony, between low [ɑ ɑ̃] and the [-low] vowels, is inactive.
- In addition, investigation of the coefficients shows that the identity predictor plays the largest role in shaping these data.

	Predictor (all binary)	Type of harmony	Coefficient
a.	*[αATR][-αATR]	ATR	0.40
b.	*[αback][-αback]	Backness	0.26
c.	*[αhigh, -back, -low][-αhigh, -back, -low]	Height amongst front vowels	0.53
e.	*[αnasal][-αnasal]	Nasal	0.63
f.	$*[+syllabic]_i[+syllabic]_j$	Identity	1.04

- ***Conclusion:*** The fact that the identity predictor is significant tells us that the identity preference in Ngbaka vowels cannot be explained entirely by appealing to interacting processes of harmony; the identity predictor plays an independent role.