AN AREAL TYPOLOGY OF NASAL VOWELS AND THE "ABSENCE" OF NASAL CONSONANTS IN NORTHERN SUB-SAHARAN AFRICA

Dmitry Idiatov & Mark Van de Velde

LLACAN (CNRS – INALCO)

dmitry.idiatov@cnrs.fr

mark.vandevelde@cnrs.fr

Langage, Langues et Cultures d'Afrique

OBJECTIVES, PRINCIPLES & METHODOLOGY

- Look for **interesting correlations** in the distribution of values of various linguistic features **in space**

- Try to find **plausible explanations** in terms of **scenarios** which would imply concrete mechanisms of linguistic change (also using data from other disciplines)

- Explanations are fundamentally **diachronic**

  "a theory of why languages are the way they are is fundamentally a theory of language change…" (Dryer 2006:56).

- Following the **methodology** developed in:

  Idiatov, Dmitry & Mark L.O. Van de Velde. 2021. The lexical distribution of labial-velar stops is a window into the linguistic prehistory of Northern Sub-Saharan Africa. *Language* 97(1). 72–107. URL

  Idiatov, Dmitry, Guillaume Segerer & Mark L.O. Van de Velde. 2021. Areal patterns of noun/verb ratios in Sub-Saharan Africa. Paper presented at the Workshop "West-central African linguistic history between Macro-Sudan Belt and Niger-Congo: commemorating the 100th anniversary of the Berlin professorship for African languages and the legacy of Diedrich Westermann", Berlin, Germany. URL

PREFERRED APPROACH: CONCEPTUALLY

- bottom-up
- big data
- garbage in, garbage out
- let the data speak for themselves (☹ binning)
- non-binary
- spell out the rules first

PREFERRED APPROACH: PRACTICALLY

- Use the **databases that exist** to harvest the data (depending on the feature of interest: **RefLex**, Phoible, ALFA, Geonames…)
- **Enrich** the harvested data with manually collected data if need be
- **Clean** and **format** the data given research questions and hypotheses and your theoretical assumptions
- Visualize the data **with different visualization methods** to confirm that the results are **qualitatively robust**

VISUALIZATIONS: SPATIAL ANALYSIS

- **deterministic** methods
  - **spatial interpolation by IDW** (inverse distance weighting): exact, finer structure
  - **spatial interpolation by Kernel smoothing** : inexact, general trends
- **statistic** (non-deterministic) methods, such as
  - GAM (generalized additive modeling)
  - GAMM (+ mixed)

- **Advantages** over deterministic methods:
  - a non-deterministic model that describes **a distribution of possible outcomes**
  - **more stable** to variations in the quantity and quality of the data
  - provides **quantified results**
  - comes with **coefficients** that allow for a more objective evaluation of the visualizations
  - can help to **discover patterns** in the data

- What is GAM?: an extension of multiple regression that provides flexible tools for modeling complex interactions describing wiggly surfaces
  - **regression**
  - wiggly surfaces
  - thin-plate splines
- A powerful tool, but still with some **limitations**
  - type of the distribution of the data (especially, non-Gaussian distributions)
  - Abrupt changes of the dependent value

LABIAL-VELARS

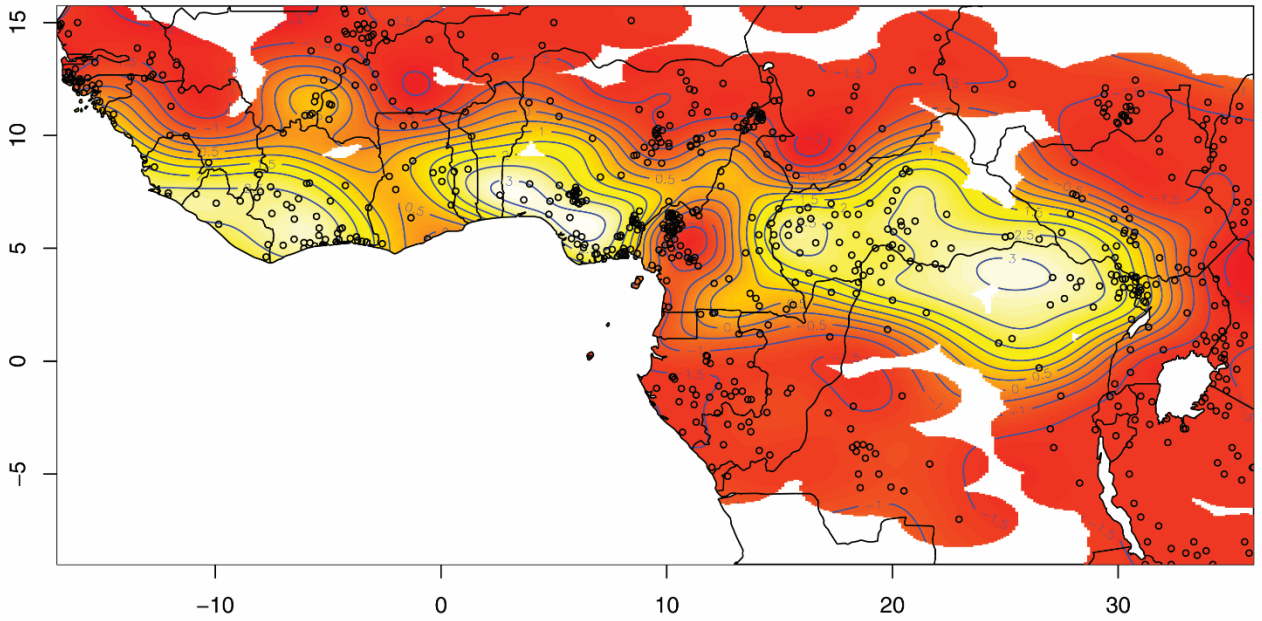
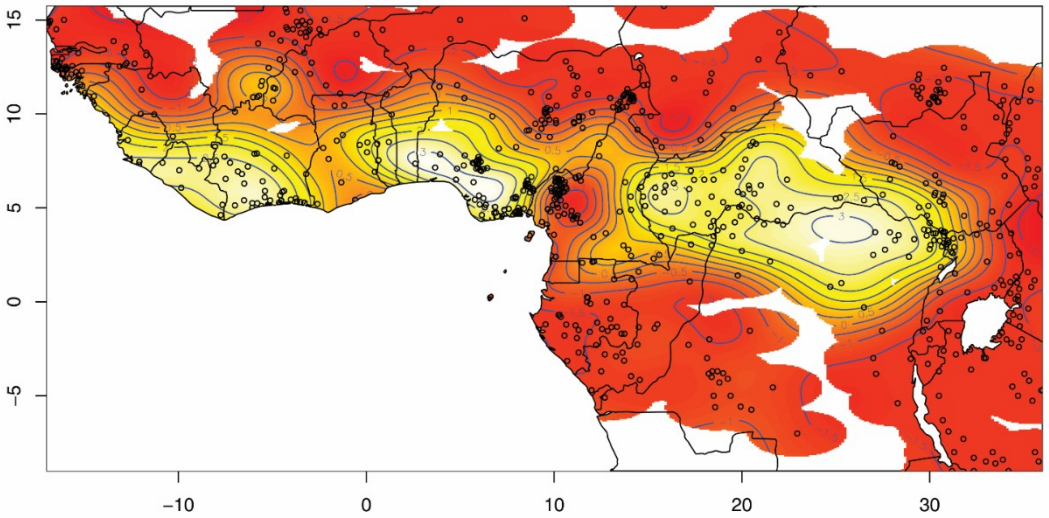
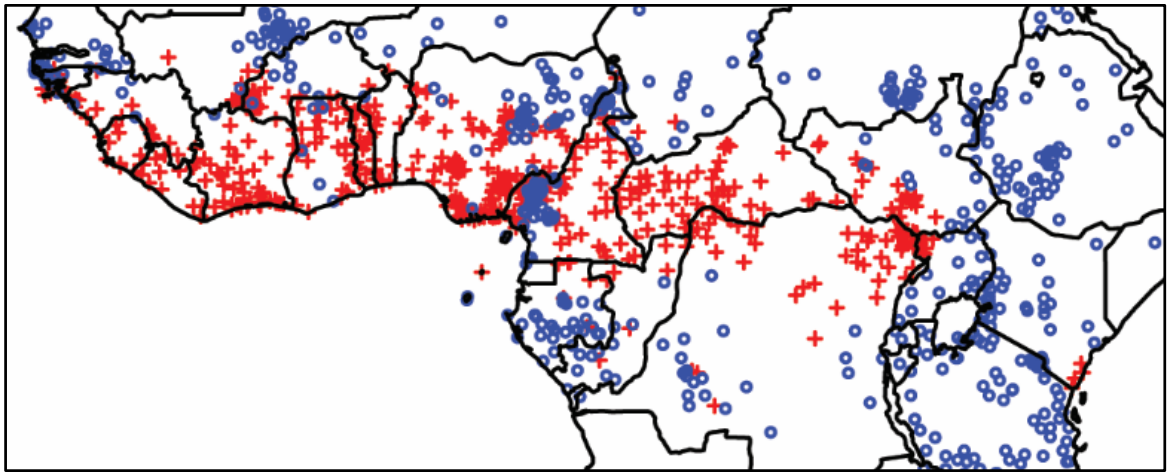


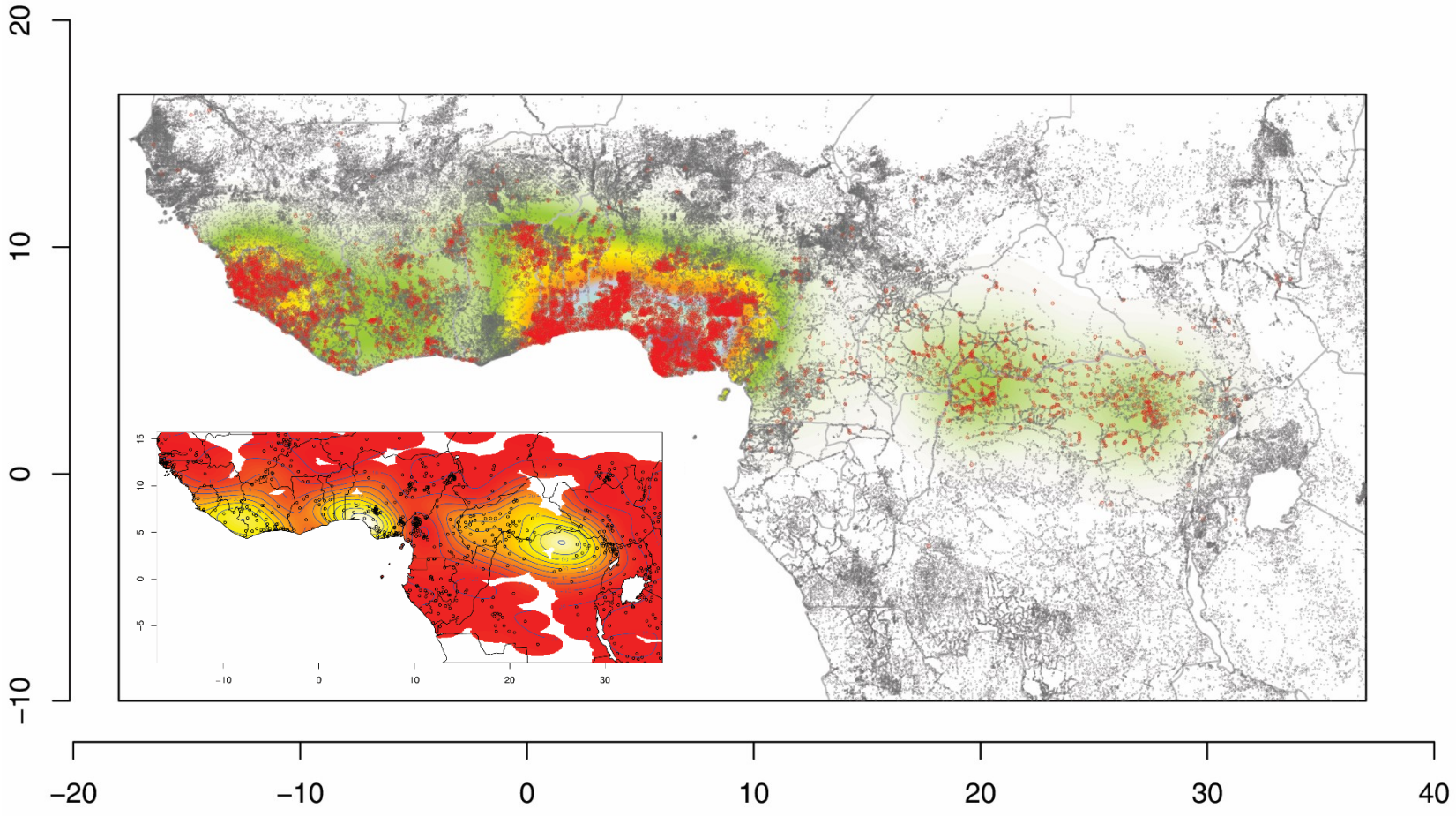
FIGURE 9 from Idiatov & Van de Velde (2021): The heat map color scheme contour plot of the GAM regression surface of the log-transformed (after scaling up by 0.83) F_{LV} frequencies (including the languages without LV stops) as a function of the combination of longitude and latitude using thin-plate regression splines. The model summary: k = 18 (k-index = 1, p-value = 0.53, k′ = 323), family = Gaussian, edf = 108.1, deviance explained = 85.80%, AIC = 1764, intercept log-transformed (after scaling up by 0.83) F_{LV} = 1.54837, $p < .001$.

STATISTIC VISUALIZATION: GAM





- **Cross-validation** with other types of data



LV STOPS: MAIN FINDINGDS

- Languages with higher lexical frequencies of LV stops are grouped into three areal hotbeds
- Languages with LV vary significantly with respect to the status of LV in their phonologies and lexicons
- In many of the languages with LV stops, they have a much lower lexical frequency than average consonant phonemes
- LV stops have a skewed lexical distribution, both phonotactically (stem-initial position) and semantically (expressive vocabulary)

LV STOPS: HISTORICAL INTERPRETATION

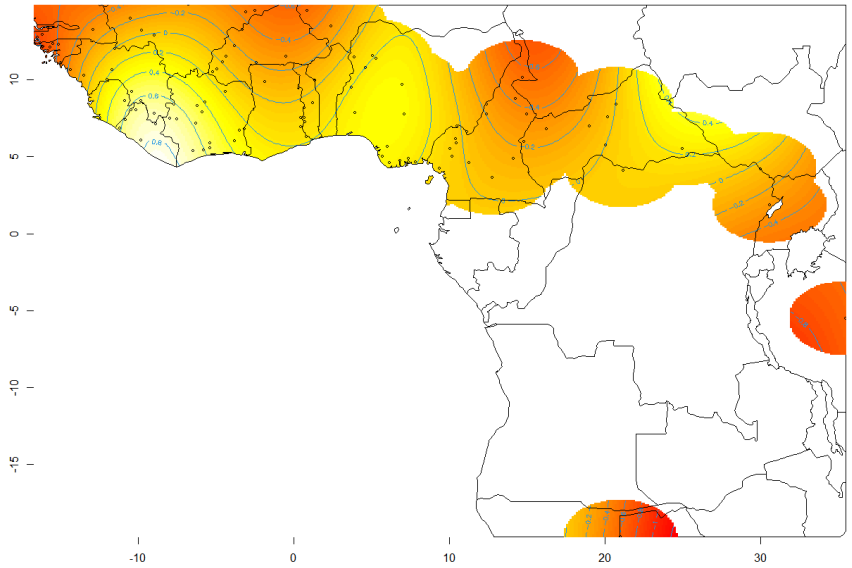
- LV stops are a substrate feature and the three hotbeds are areas of retention and refuge zones.
- LV stops are retentions from an areal point of view, but innovations from a genealogical point of view in the great majority of African languages that have them today.
- Detailed hypotheses regarding prehistoric migration patterns of Niger-Congo speaking populations
- Adjusted and refined the scenarios for the Bantu expansion.
- C-emphasis prosody as the primary force driving the emergence, spread, and intra-linguistic distribution of LV stops

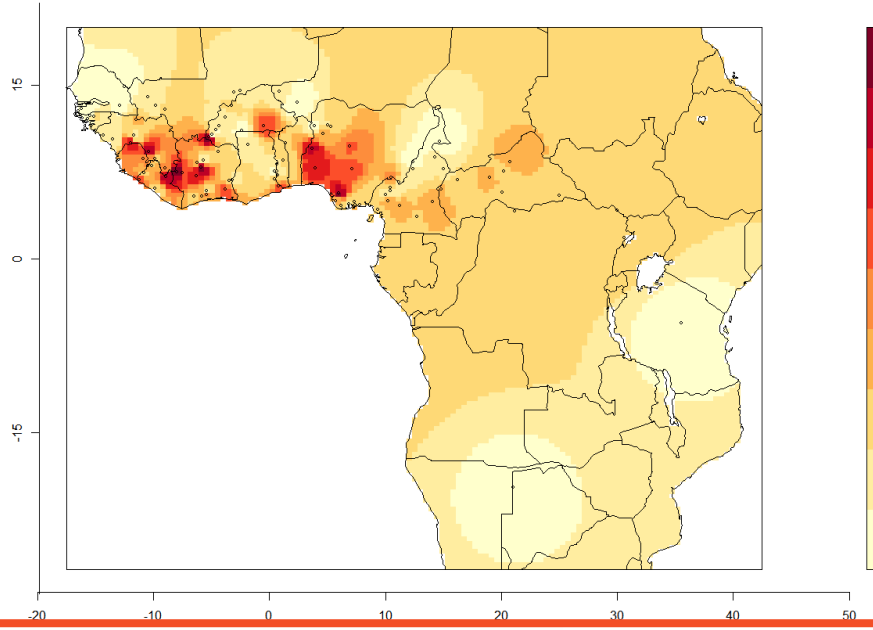
NOUN/VERB RATIOS

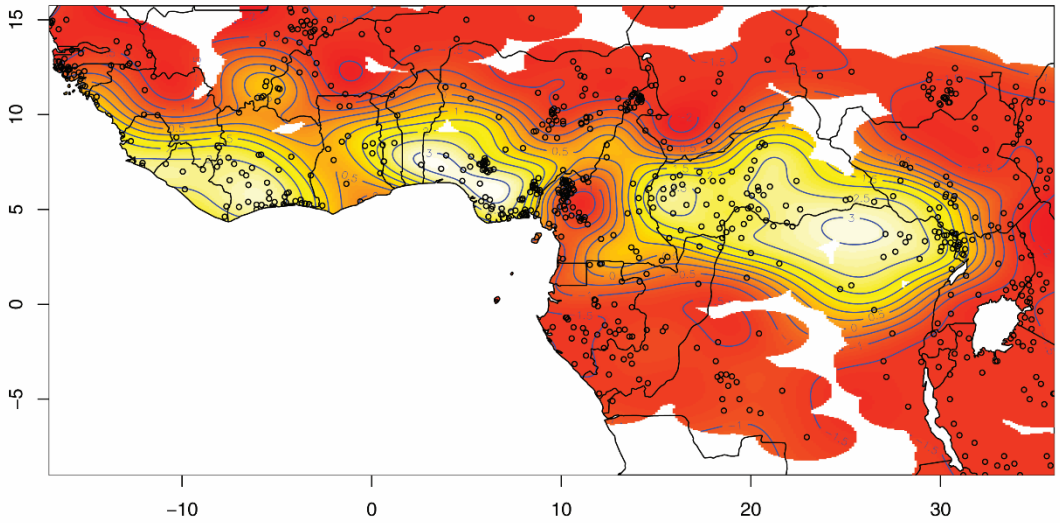
SUBSTRATE EFFECTS IN MORPHOSYNTAX: N/V RATIOS

- The same methodology can be applied to morphosyntactic patterns
- N/V ratios in Sub-Saharan languages show striking, areally conditioned differences that reflect substrate effects (Idiatov, Segerer & Van de Velde 2021)

N/V RATIOS PRELIMINARY RESULTS: 1H2L vs LV HOTBEDS







N/V RATIOS PRELIMINARY RESULTS

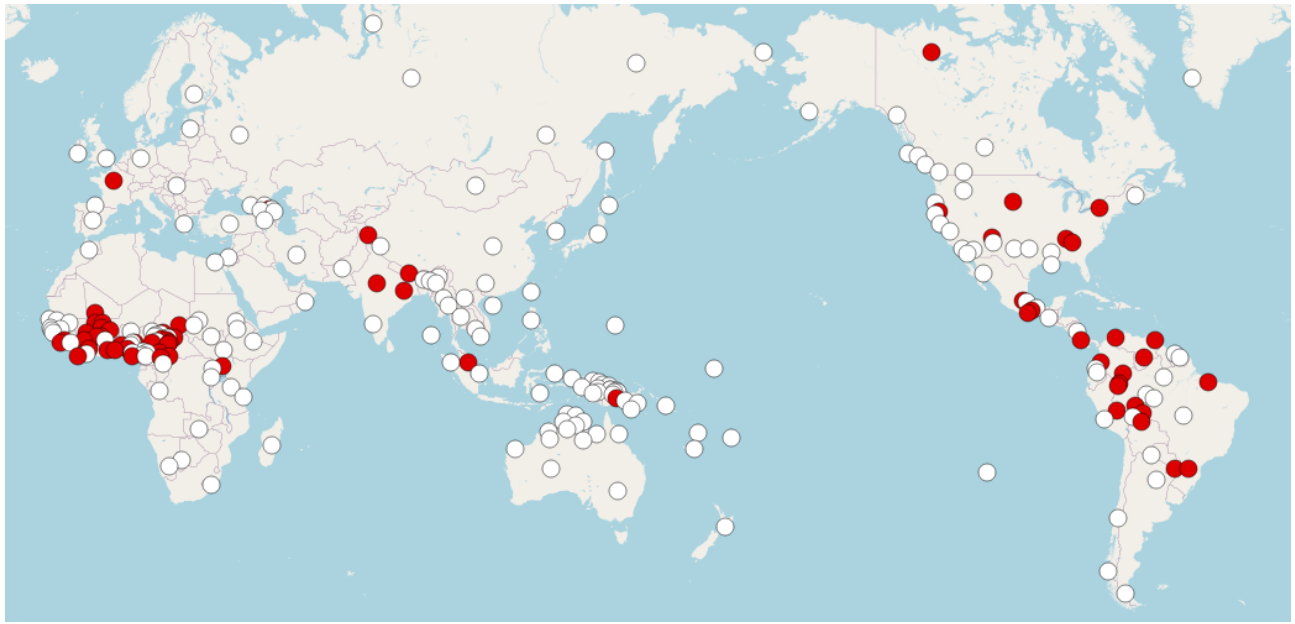
Preliminary results with respect to N/V ratios in (N)SSA:

- Languages with few verbs (high N/V ratios) are concentrated in two areal hotbeds
- These two hotbeds largely **coincide with** the Lower and Upper Guinea hotbeds of high lexical frequency of **LV stops**
- The Ubangi Basin hotbed, in contrast, does not clearly correspond to an area with a high N/V ratio

NASAL VOWELS

NASAL VOWELS: NSSA VS WORLD

- Contrastive nasal vowels are **particularly common in NSSA** when compared to the rest of the world.

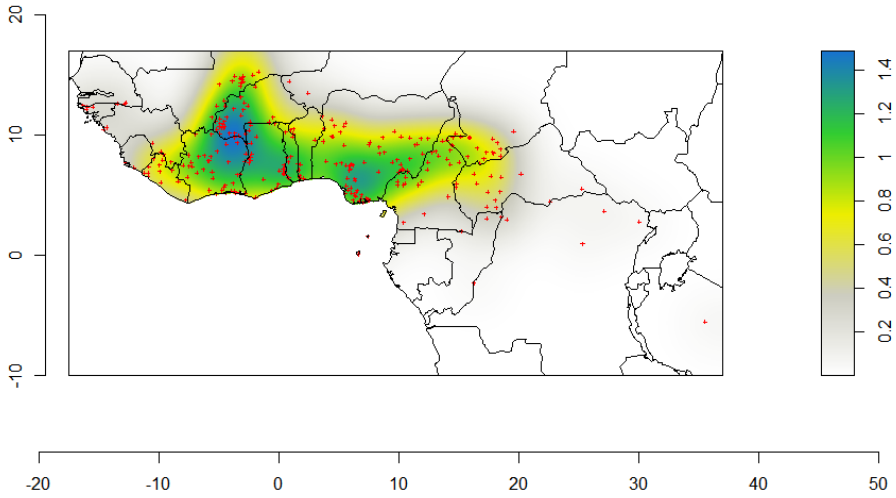


Hajek (2013) in WALS feature 10A "Vowel nasalization"

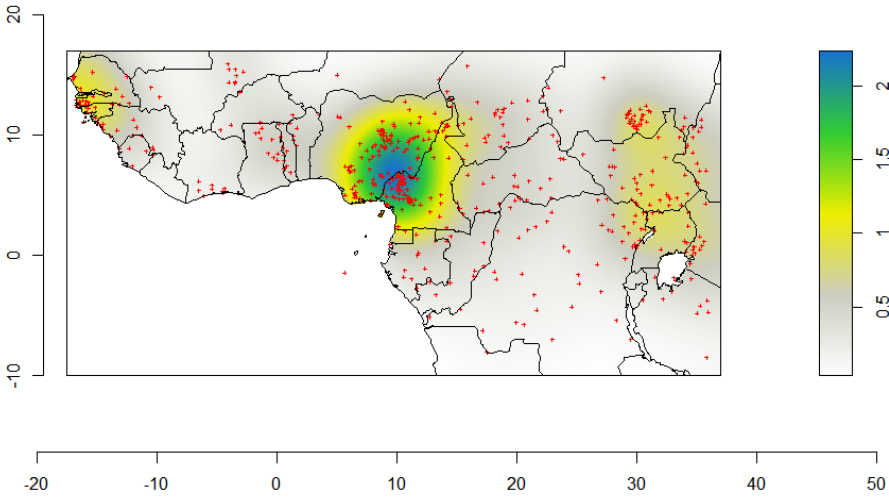
- Considered as one of its **defining areal features** (Clements & Rialland 2008; Hajek 2013; Rolle 2013)

NASAL VOWELS: NSSA

NSSA languages **with** contrastive nasal vowels (294)



NSSA languages **without** contrastive nasal vowels (515)



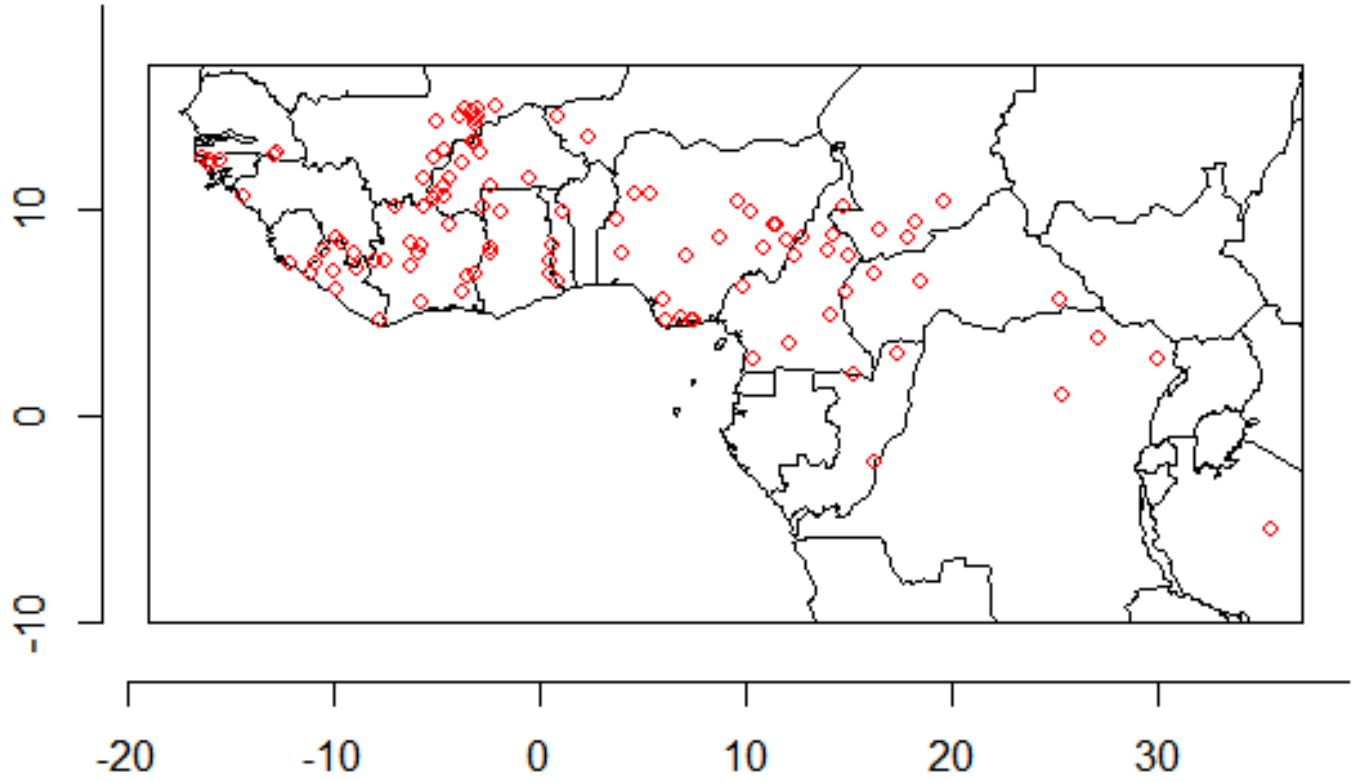
- Based on: ALFA (Rolle et al. 2020), RefLex (Segerer & Flavier 2011-2025)
  - A few conflicts
  - Not all RefLex sources taken into consideration
  - ☹ languages with nasal vowels only in borrowed lexicon
  - ☹ languages with nasal vowels only in onomatopoeia and ideophones

LEXICAL FREQUENCY OF NASAL VOWELS

- The lexical frequency data come from RefLex (www.reflex.cnrs.fr)
- RefLex has 2196 sources for more than 1100 languages, but the source are of very uneven quality
- Selection procedure for sources:
  - Limited to NSSA: longitude interval $[-18°, 36°]$, latitude interval $[-9°, 16°]$
  - Sources $>$ 400 entries (cf. Dockum & Bowern 2019)
  - Sources published after 1900
  - Remove comparative wordlists (TLS, BCCW, ALGAB, Koelle)
  - One source per language
  - Manual quality checkup

LEXICAL FREQUENCY OF NASAL VOWELS

- 113 languages with data on lexical frequency of nasal vowels



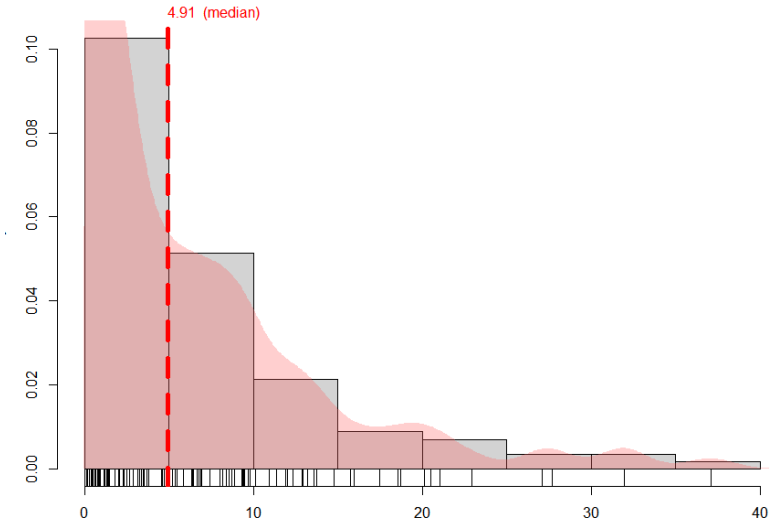
LEXICAL FREQUENCY OF NASAL VOWELS

$$F_{NasV} = Tokens_{NasV} / Tokens_{(NasVowels + OralVowels)} * 100\%$$

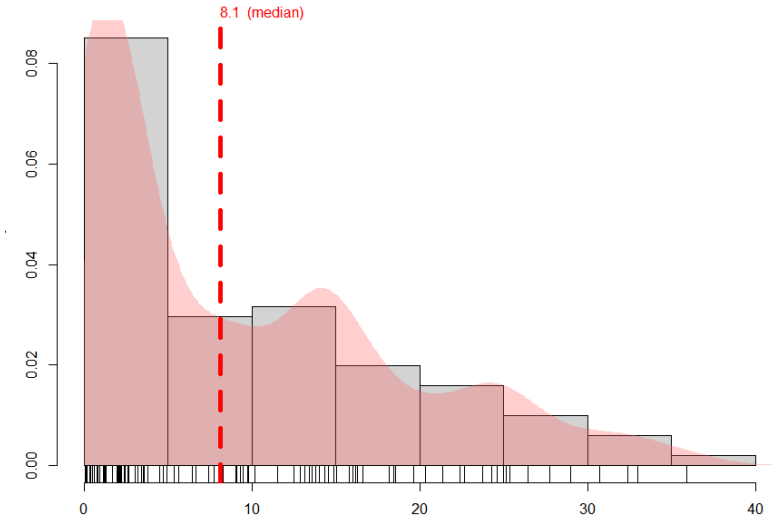
- Two kinds of lexical frequency estimation (in percentages):
  - **FreqTokens**: The token frequency of nasal vowels in the source as a whole.
  - **Freq1stSylVerbs**: The token frequency of nasal vowels in the first syllable of verbs which begin with a simple oral plosive or fricative C (that is, no nasals, no implosives, no laterals, no rhotics, no approximants, no consonant clusters) or a vowel
- The **overall results** for the 2 types of frequency estimations are **very similar**
- For languages, for which we have several sources, the estimations based on different sources strongly tend to **agree**

LEXICAL FREQUENCY OF NASAL VOWELS

- Nasal vowels tend to be **rare** in languages that have them.



Probability density for FreqTokens

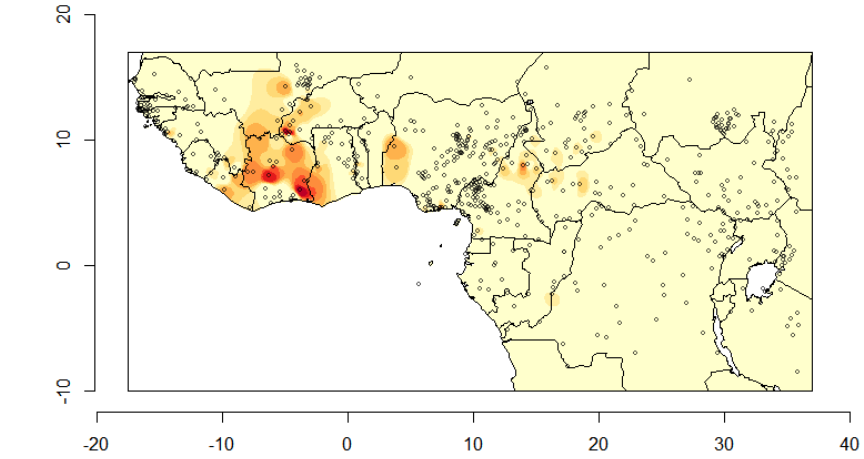


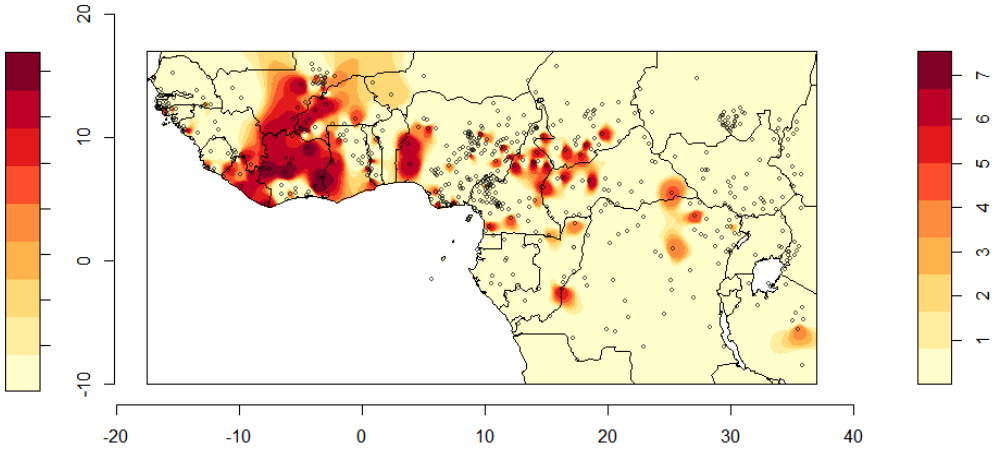
Probability density for Freq1stSylVerbs

- Compare labial-velars…
- Log-transformation to zoom in on lower frequency values

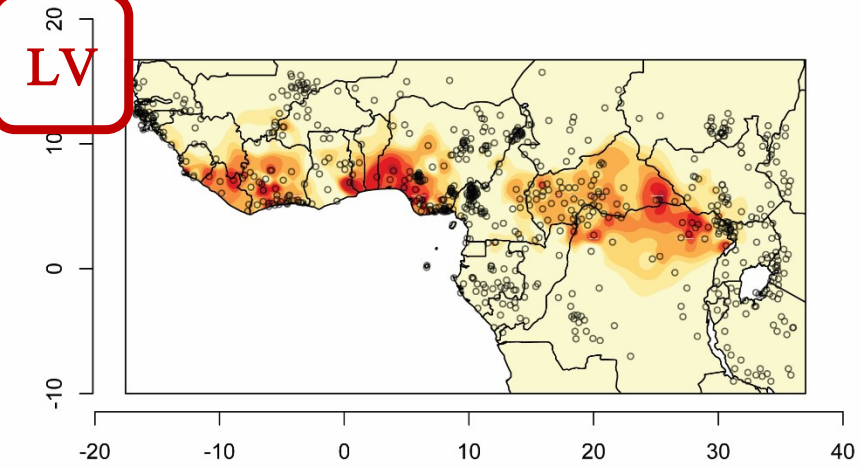
LEXICAL FREQUENCY OF NASAL VOWELS

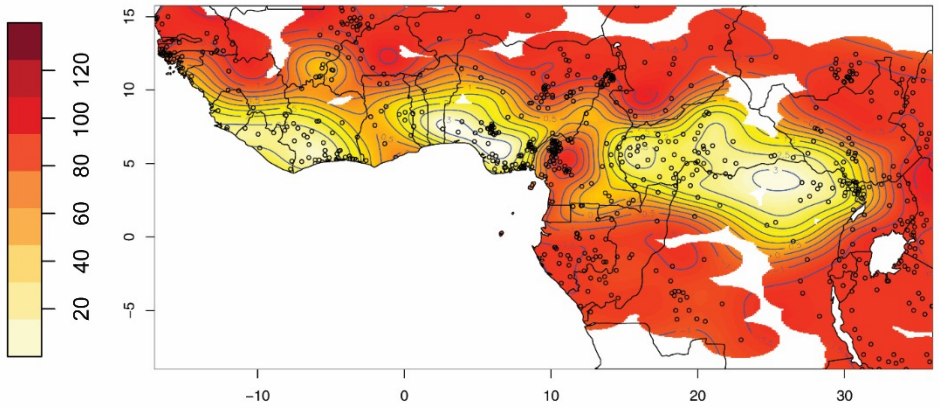
- IDW of FreqTokens: base & log-transformed





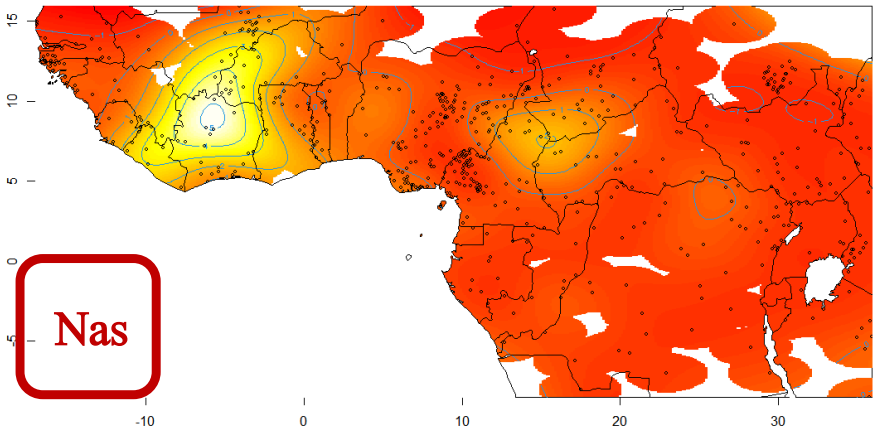
LV

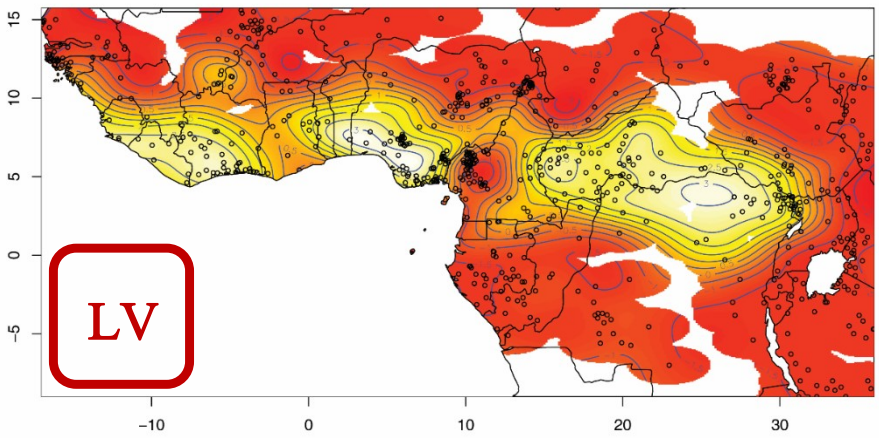


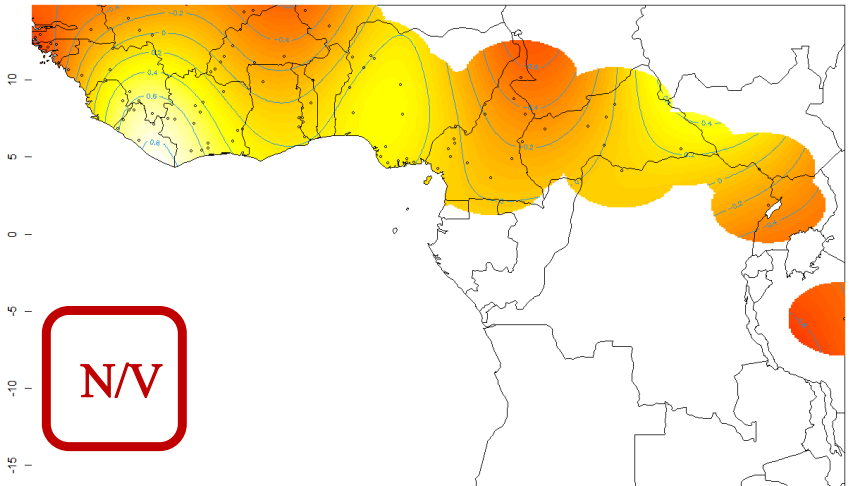


LEXICAL FREQUENCY OF NASAL VOWELS

- GAM model of FreqToqens_{LOG} vs LV_{LOG} vs N/V ratios







NASAL VOWELS: SEMANTIC SKEWING

- In languages with low lexical frequencies of nasal vowels, these often show a distribution that is **semantically skewed**

  ✋ Somewhat like labial-velars… (cf. Idiatov & Van de Velde 2021)

  - borrowings

    Bedik (North Atlantic) *lãsɛ̀t* 'razor blade' (< FR), Pichi (Creole) grɑ̃frɛ̀r 'older brother' (<FR), Vai (Mande) *pɑ̃ĩ̀* 'pint'

  - onomatopoeia

    Basari (North Atlantic) *xɛ̃ xɛ̃ xɛ̃* 'cry of a kind of bird'

  - ideophonic and expressive vocabulary

    Lega-Beya (Bantu) *kɑ̃̀kɑ̃̀kɑ̃̀* 'emphatic insistence', Bullom (Mel) *hɑ́ɑ́ɑ́* 'deep, far, long', Furu (Bongo-Bagirmi) *ʉ́ʉ́* 'long time ago', Vai (Mande) *kpɑ́* 'firmly', *dɛ́ĩ́dɛ̀ĩ̀* 'epilepsy', Looma *vɑ́ɑ́vɑ́ɑ́* 'slowly'

  - interjections (often, 'yes' and 'no')

    Aghem (Bantoid) *ɔ̀ɔ̃* 'yes', Ndut (North Atlanic) *ĩ ~ ĩ̀ĩ̀,* Mamvu (Membi-Mangbutu-Efe) *ĩhĩ* 'expression of rebuke', Looma (Mande) *ʉ̀ʉ̀* 'yikes', *ɛ̀ɛ̀* 'hmm. (hesitation)'

NASAL VOWELS: SEMANTIC SKEWING

- In languages with low lexical frequencies of nasal vowels, these often show a distribution that is **semantically skewed**

  ✋ Somewhat like labial-velars… (cf. Idiatov & Van de Velde 2021)

  - species terms

    Vai (Mande) *vɔ̃̀ɔ̃́vɔ̃̀ɔ̃̀* 'hornbill', *lóã́* 'kind of tree', *kpãã̀kɛ̀sì* 'wasp'

  - specialist vocabulary

    Vai (Mande) *tòã̀* 'smithy', *kpɛ́ɛ́sì* 'remove (palm nuts from among thorns of cluster)'

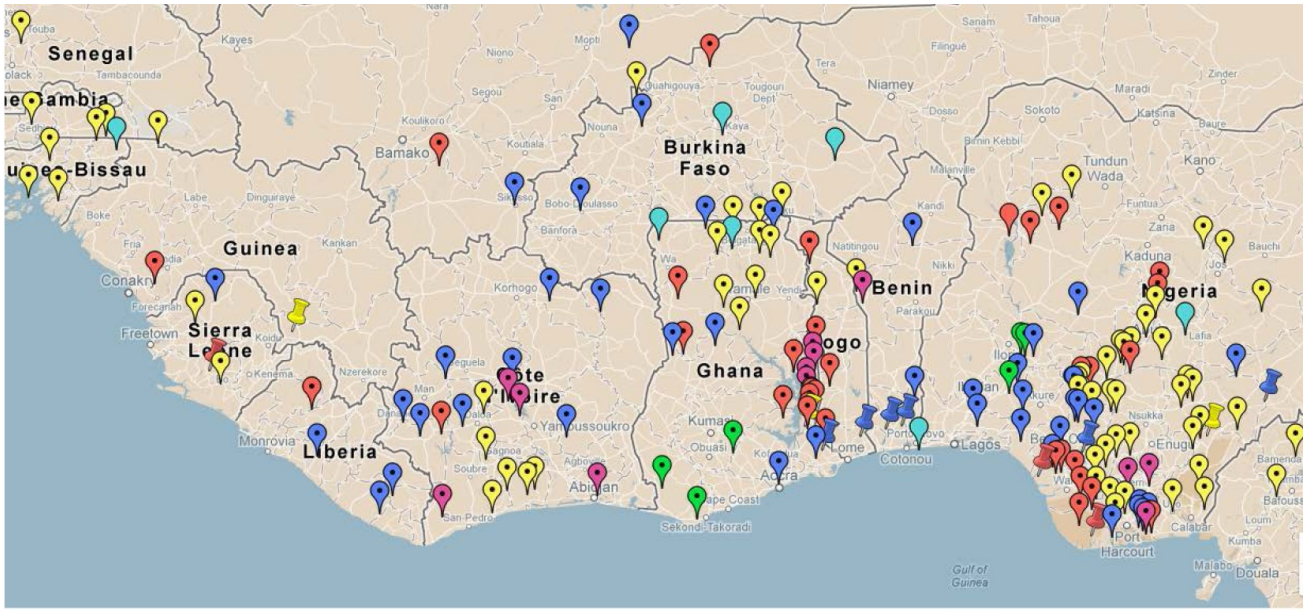
  Mende (Mande; Innes 1968):

  - 311 out of 7937 entries (= 3,9%) have a nasal vowel
  - 162 (= 52%) of the entries with a nasal vowel are ideophones
  - Only 914 (= 11,5%) out of 7937 entries are ideophones.

NASAL VOWELS AND CONSONANTS

MID-HIGH NASAL VOWELS

- Restrictions on **mid-high nasal vowels** (Hyman 1972; Rolle 2013)
  - **/ẽ, õ/** are frequently **absent** in the inventories of nasal vowels
    - This is phonetically natural, but still remarkable cross-linguistically (Rolle 2013)



MID-HIGH NASAL VOWELS

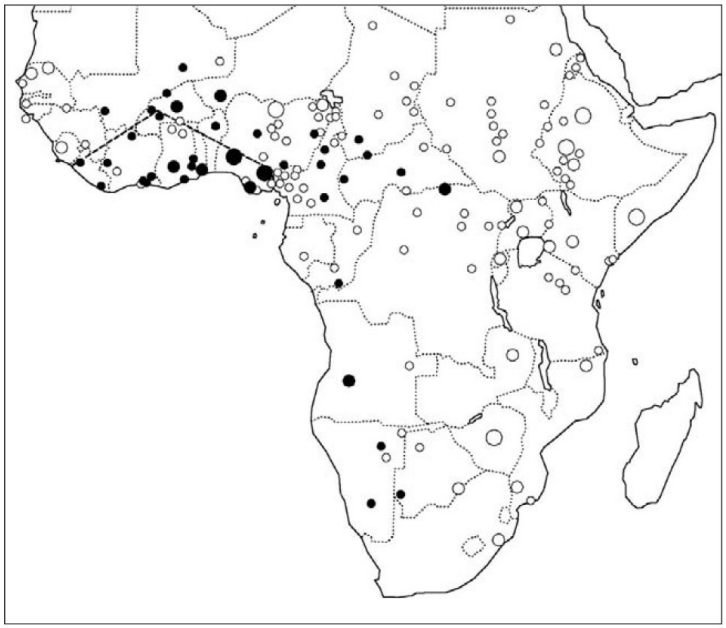
- Restrictions on **mid-high nasal vowels** (Hyman 1972; Rolle 2013)
  - A frequent **phonotactic restriction** (or dispreference) on sequences: ***[nẽ ~ ne, nõ ~ no]** and **[mẽ ~ me, mõ ~ mo]**
    - Originally, with respect to the Kwa/Benue-Congo languages
    - But it is more widespread and may apply to other nasal consonants too:
      - Bambara (Mande), with /õ, ẽ/ and both NV and NṼ (Dumestre 2011 with 23170 entries):

        mõ (1), nõ (1), nẽ (1) ; *mẽ, *ɲẽ, *ɲo
      - Grebo (Kru), no /õ, ẽ/ and (almost) only NṼ (Innes 1967 with 6917 entries):

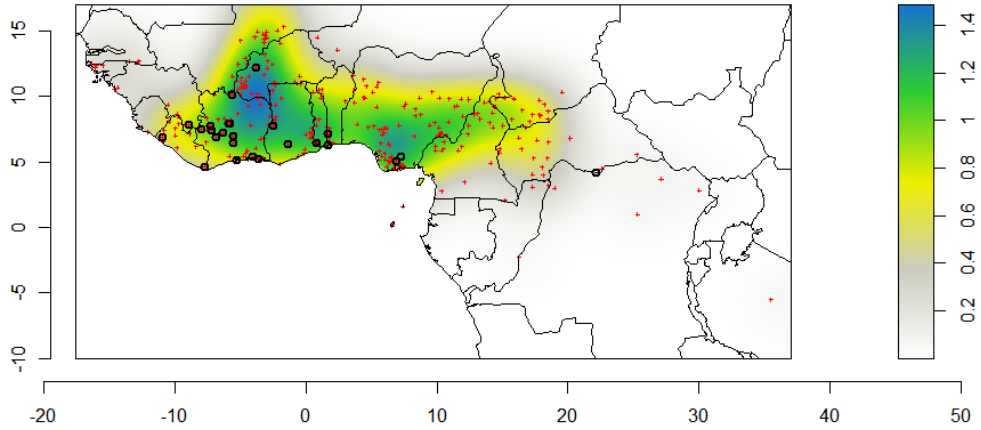
        mo (1), no (1), ɲe (1) ; *me, *ne, *ɲo, *N_{other} + o/e

NO CONTRASTIVE NASALS

- The possibility to analyze various languages as **lacking contrastive nasal consonants** (cf. Bearth 1992; Bole-Richard 1985; Clements & Rialland 2008; Hyman 1972; Ladefoged 1964; Schachter & Fromkin 1968)



Map 3.3 Distribution of contrastive nasal vowels in a sample of 150 African languages. The area enclosed in dashes contains languages reported to lack distinctive nasal consonants



Clements & Rialland (2008:46)

NO CONTRASTIVE NASALS

- The possibility to analyze various languages as **lacking contrastive nasal consonants** (cf. Bearth 1992; Bole-Richard 1985; Clements & Rialland 2008; Hyman 1972; Ladefoged 1964; Schachter & Fromkin 1968)

> "Such languages typically have an oral vs. nasal contrast in vowels, and two sets of consonants. Members of set 1 are usually all obstruents and are realized as oral regardless of whether the following vowel is oral or nasal. Members of set 2 are usually non-obstruents, and are realized as oral sounds before oral vowels and as nasal or nasalized sounds before nasal vowels."
>
> Clements & Rialland (2008:46-47)

Ikwere (Igboid)

(1) before oral vowels (set 2a)		before nasal vowels (set 2b)	
áḅ á	'paint'	ámà̰	'matchet'
á'ḅ á	'companionship'	à'mà̰	'path, road'
ɔ̀-lʊ́	'to marry'	ɔ̀-nʊ̰́	'to hear'
érú	'mushroom'	ɛ́rʊ̰́	'work'
à-yá	'to return'	áỹâ̰	'eye'

NO CONTRASTIVE NASALS

- The possibility to analyze various languages as **lacking contrastive nasal consonants** (cf. Bearth 1992; Bole-Richard 1985; Clements & Rialland 2008; Hyman 1972; Ladefoged 1964; Schachter & Fromkin 1968)

  - "[M]any West African nasal systems can be ranged along a continuum in regard to the plausibility of a ''no-nasal'' analysis" (Clements & Rialland 2008:49)

    …and in our view, it largely remains a (somewhat misleading) idealization of more complex phonological realities of the languages in question (see also Bearth 1992; Fromkin 1977).

NO CONTRASTIVE NASALS

- Clements & Rialland (2008:47) cite **25 languages** as "reported to **lack distinctive nasal consonants**".

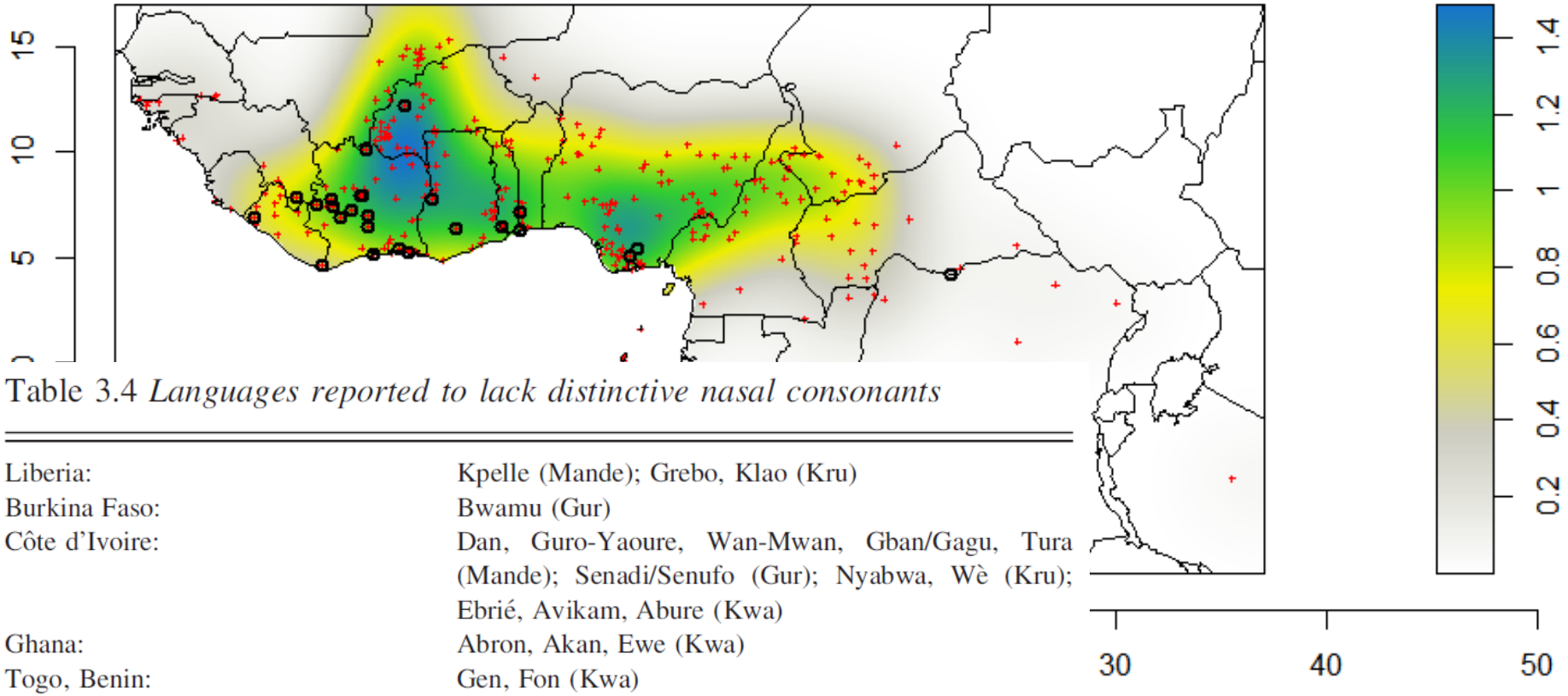


Table 3.4 *Languages reported to lack distinctive nasal consonants*

Liberia:	Kpelle (Mande); Grebo, Klao (Kru)
Burkina Faso:	Bwamu (Gur)
Côte d'Ivoire:	Dan, Guro-Yaoure, Wan-Mwan, Gban/Gagu, Tura (Mande); Senadi/Senufo (Gur); Nyabwa, Wè (Kru); Ebrié, Avikam, Abure (Kwa)
Ghana:	Abron, Akan, Ewe (Kwa)
Togo, Benin:	Gen, Fon (Kwa)
Nigeria:	Mbaise Igbo, Ikwere (Igboid)
CAR:	Yakoma (Ubangi)

NO CONTRASTIVE NASALS

- Kpelle (Konoshenko 2017 among others)
  - It does have /ŋ/, so the feature [+nasal] is needed for its consonants anyway
  - NṼ vs NV (the nasalisation of the vowel is predictable only when we know the morphology)

    [(ń)nɑ̃̂ŋ] ‘my father’ vs. [(ń)nâŋ] ‘to make me jump’ (the nasalisation of the vowel is predictable only when we know the morphology
  - LṼ, BṼ

    [lónó ~ lɔ̃́nɔ̃́] ‘conversation’

    [ɓénéŋ ~ ɓɛ̃́nɛ̃́ŋ ~ mɛ̃́nɛ̃́ŋ] ‘fonio’

    [ɓɔ̃́mɔ̃́] ‘wax]

NO CONTRASTIVE NASALS

- Tura (Bearth 1971, 1992 ; own data)
  - It does have /ŋ/, so the feature [+nasal] is needed for its consonants anyway

    ✋ The same applies to all other Southern Mande languages on that list: Dan, Guro, Yaure, Mwan, Gban
  - NṼ vs NV (the nasalisation of the vowel is predictable only when we know the morphology)

    [àmmā́] 'hear them' vs. [àmmà] 'of them'
  - At least a few words consistenly [NV] (with a mid-high vowel…):

    [mɔ̀] PL allormorph (lexically conditioned)

    [-nɔ́] 'every-', as in [mɛ́nɔ́] 'everyone'.

NO CONTRASTIVE NASALS

- Grebo (Innes 1966, 1967)
  - At least a few words consistently **[NV]** (with a mid-high vowel…):

    [mőbò] ‘kind of grass’

    [nòbò] ‘central stalk on which the fruit of palm trees grows’

    [ŋ̀bè] ‘a kind of antelope’

NO CONTRASTIVE NASALS

- Ikwere (Osu & Clements 2009)
  - **V > Ṽ / n-** 'PROG' (with a mid-high vowel…), resulting in [NṼ] where the source of the nasalization is not the vowel.

    [èrî] 'eat' > [n-ẽ̀rî] PROG = eat

NO CONTRASTIVE NASALS

- A bet: If any of these languages has N-final words and V-initial words, such a word-initial V would not be nasalized after a word-final N

"NO CONTRASTIVE NASALS": THE SOURCE

✋ All sequences below are **tautomorphemic** (or at least **word-internal**) and consequently the changes are **morphonological**

Stage 0: NV, DV

Stage 1: NV, NṼ, DV, DṼ

Nasal vowels emerge through a number of processes: *CVNV > CNV > CṼ (Hyman 1972), *CVNCV ~ *CVNV > CṼNṼ > CṼṼ > CṼ (Williamson 1973; Welmers 1976) ; *CVN > CṼ

Stage 2A: (articulatory-driven) perseveratory nasalization: NV > NṼ

Stage 2B: (perceptually-driven) anticipatory nasalization DṼ > NṼ

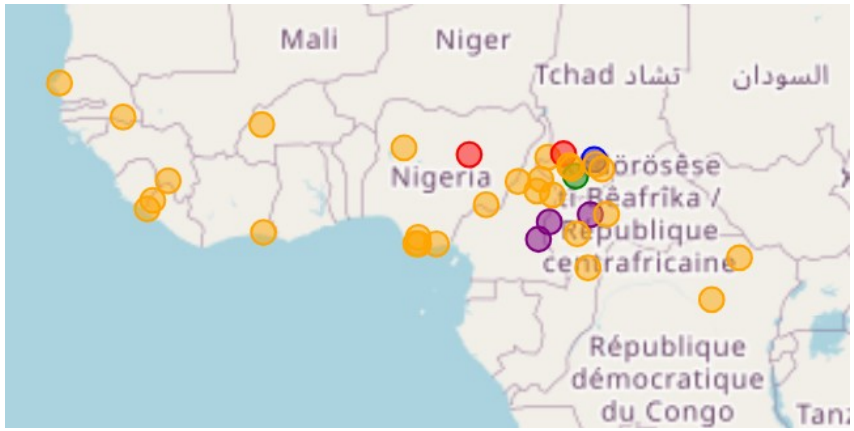
affecting implosives, approximants and subsequently laterals and rhotics

Stage 4: NṼ, DV

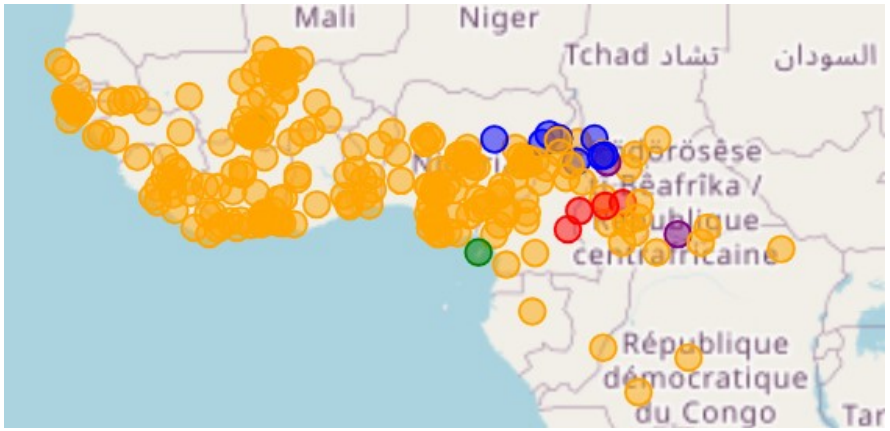
☞ It is **the combination of its pre-conditions and subsequent changes** that makes this pattern **rare cross-linguistically**.

NO CONTRASTIVE NASALS

ƁṼ: 38 languages & 142 entries



L/RṼ: 328 languages & 6761 entries



- There is nothing in the articulation of ƁṼ that would make it particularly difficult to pronounce.
- It is probably the lack or low intensity of the burst at the release of implosives that makes them particularly prone to **perceptual confusion** in the context of a tautosyllabic nasal vowel.