Liquid consonants and onset sonority in Dogon languages

Michael Dow
Université de Montréal

50 ans de linguistique à l'UQAM

~~March 28, 2020~~
April 24, 2021

Introduction

Goals

- Looking at lexical patterns in Dogon languages in the hopes of explaining the exceptional behaviour of liquid consonants in many of these languages. For example:
  - Dissimilation-like behaviour affecting both affix consonants (e.g., Ben Tey /pi**l**e-**l**e/ ⟶ [pi**l**e-**r**e] 'white-INCHOATIVE'; Heath 2015a) and root consonants (e.g., Nanga /kɔ**r**i-**r**i/ ⟶ [kɔ**ll**i-**r**i] 'hook-REVERSIVE'; Heath 2016)
  - Liquid mutations (e.g., Toro Tegu [bɛ**r**u] 'near' ~ [bɛ**l**a] 'near-INCH'; Heath 2015b)
  - A 'flip-frop' (e.g., Ben Tey /ɔ**r**u-**l**i/ ⟶ [ɔ**l**i-**r**i] 'moist-INCH'; Dow et al. 2017)
- Emergent preference for onset profile [l. . . r], usually between second and third syllables at a morpheme boundary.

Existing analyses

- Prosody seems important in capturing these facts:
  - Toro Tegu (Heath, 2015b): Unstressed vowel deletion leads to rhotic deletion /bɛru-la/ ⟶ |bɛr-la| ⟶ [bɛla] 'far-INCH'
  - Ben Tey (Dow et al., 2017): Embedded feet (superfeet) in trisyllabic words fortify second position
- Unfortunately, little is written about Dogon prosody, besides trochaic (or "second-syllable weakness") analysis (Heath, 2008; McPherson, 2013)

Today's talk

- Lexical statistics from a corpus of 12 Dogon languages (nearly 29,000 monomorphs): Which onset consonants do we observe, according to syllable position and word length?
- Brief comparison with French
- Groundwork for future analysis: sonority licensing according to strength of position

Background

Dogon languages

- Family of ~20 Niger-Congo languages mostly spoken in Mali, Mopti province

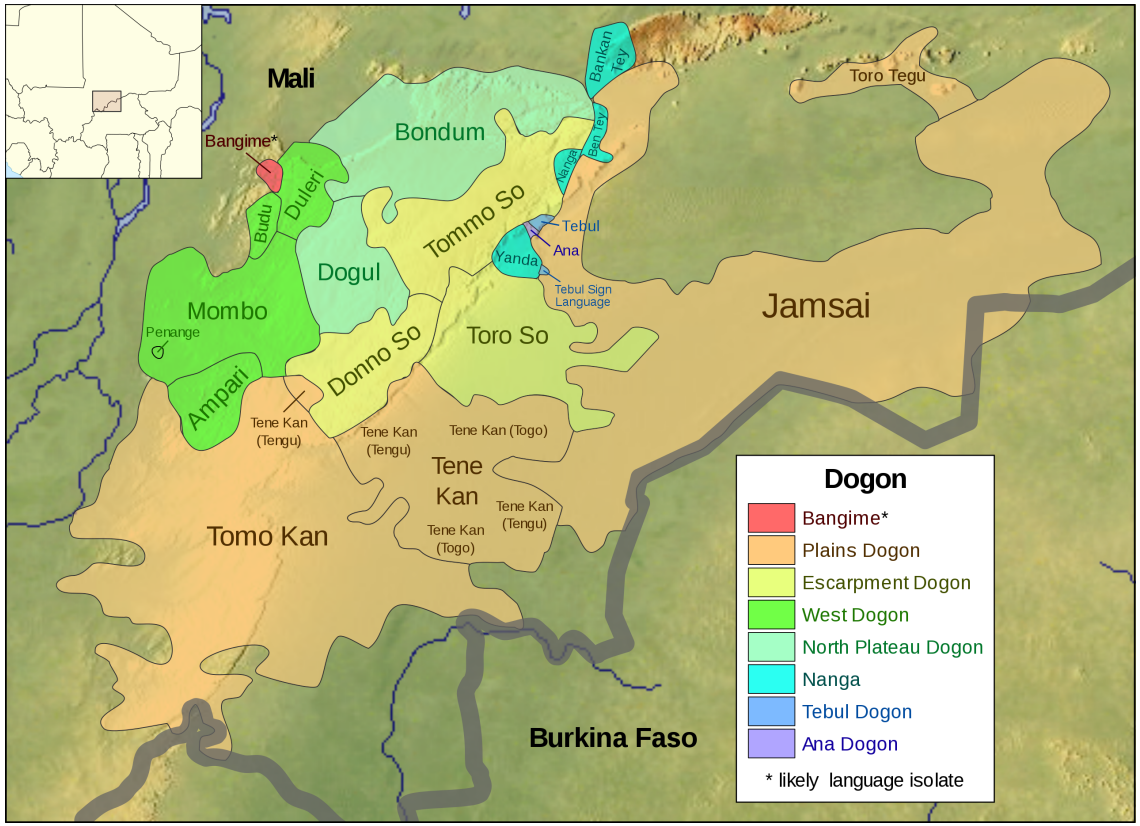

Fig. 1. Map of Dogon languages. Subgroups are not necessarily standard. Source: `https://en.wikipedia.org/wiki/Dogon_languages{#}/media/File:Map_of_the_Dogon_languages.svg`

Sonority & onsets

- Cross-linguistically, lower-sonority consonants such as stops are preferred in onset positions.
- There are perceptual as well as theoretical explanations for this tendency (e.g., Smith, 2004).
- Generally, despite these preferences, we don't see the same kind of language-internal restrictions on onsets that we see on codas, all things being equal (Rousset, 2004)

Tying in prosody

- However, we can see disparities in non-assimilative neutralization such that sounds increase in sonority in weak positions, such as coda position and foot-internal onsets (Harris, 1997)
- In terms of structure, this can generally be seen as a failure to authorize extra material in such positions, for instance, in terms of elements or the sort of *xo* theory approach to sonority used in De Lacy (2006).
- Lower-sonority onsets can even contribute weight and attract stress by virtue of their sonority (e.g., Gordon, 2005; Ryan, 2019)
- In other words, obstruents generally have more structure than sonorant consonants, and stronger positions allow for this (at the language's discretion).

Methodology

Corpus construction

- Initial data collected from the Lexicon page of the Dogon and Bangime Languages website (Moran et al., 2016).
- Preliminary processing:
  - Tones and vowel length removed
  - Clitic-affix distinction levelled (e.g., '=' replaced by '-')
  - All English or metalinguistic material removed
  - Complex words split by whitespace (e.g., Ampari [kìmɛ́-gé dɛ̀ndɛ̀] 'mushroom (with cap)' ⟶ [kimɛ-ge] and [dɛndɛ])
- Polymorphemic forms identified via '-' and removed
- Language-internal duplicates removed and variants within entries (incl. reduplication) reduced to single form (e.g., Perge Tegu [wárú wàrá] 'work in the fields' ⟶ [wara])
- Languages with fewer than 1000 forms remaining removed for a total of 28,795 forms from 12 languages

Corpus: Monomorphs

Language	**Forms**	**Language**	**Forms**
Bankan Tey	2350	Perge Tegu	2118
Ben Tey	2777	Togo Kan	1858
Jamsay	2268	Tommo So	2602
Mombo	3115	Toro Tegu	2459
Najamba-Kindige	2754	Yanda Dom	2189
Nanga	2463	Yorno So	1965

Table 1: Number of forms in database per language

Segmentation

- Consonants separated into larger classes (P: plosives & affricates, F: fricatives, N: nasals, L: laterals, R: rhotics, J: semivowels) and vowels defined as 'V'
- Coronal stops ('T') tagged separately (in case of flapping analysis)
- Generalized transcriptions created based on classes (e.g., [balaŋgal] 'pole for harnessing donkeys' $\longrightarrow$ PVLVNPVL)
- Gliding assumed in non-low V_1 + any V_2 sequences (in keeping with McPherson, 2013), otherwise "dummy" character '.' used for onsetless syllables

Syllabification

- Syllabification automatically performed on generalized transcriptions with Regular Expressions, with verification of random subsets performed throughout
- Only final consonant of (rare) clusters placed in onset, as complex onsets are disallowed in Dogon languages (with the exception of consonant + semivowel sequences, whose first member was selected).
- Problematic? Probably not. (Only 12 forms with internal obstruent + liquid sequences, 86 consonant + semivowel)

Data

- All remaining onsets were extracted with reference to left-edge and '.', along with:
  - Total number of syllables (V count)
  - Each onset's respective position (i.e., first syllable, second, and so on)
- Proportion tables for onset sonority profile (e.g., P...P, etc.) over two adjacent syllables calculated within languages, according to word size and syllable position
- This gives us, separately:
  - $\sigma 1 - \sigma 2$ profile in words containing 2, 3 and 4 syllables
  - $\sigma 2 - \sigma 3$ profile in words containing 3 and 4 syllables
  - $\sigma 3 - \sigma 4$ profile in words containing 4 syllables

Example

1.	[báúrú bàúrɔ́]	[gìwná ìsìndà]		importation
2.	bauru baurɔ	giwna isinda		tone removal
3.	bauru baurɔ	giwna	isinda	complex word separation
4.	baurɔ	—	—	competitor elimination
5.	ba.urɔ	—	.isinda	onsetless syllable resolution
6.	PV.VRV	PVJNV	.VFVNTV	sonority categorization
7.	{P, ., R}	{P, N}	{., F, T}	onset sonority extraction
8.	{a, u, ɔ}	{i, a}	{i, i, a}	vowel extraction
9.	3	2	3	syllable count

Table 2: Processing of two representative entries

Results

Guide to figures

- Languages broken out
- Onsetless syllables removed for legibility
- Earlier syllable on bottom
- Sonority increasing left to right and bottom to top within each breakout
- Shading indicates proportion of onset combination within conditions (language & length)
- For instance, 188 PVPV disyllables in Tommo So, vs. 68 PVRV

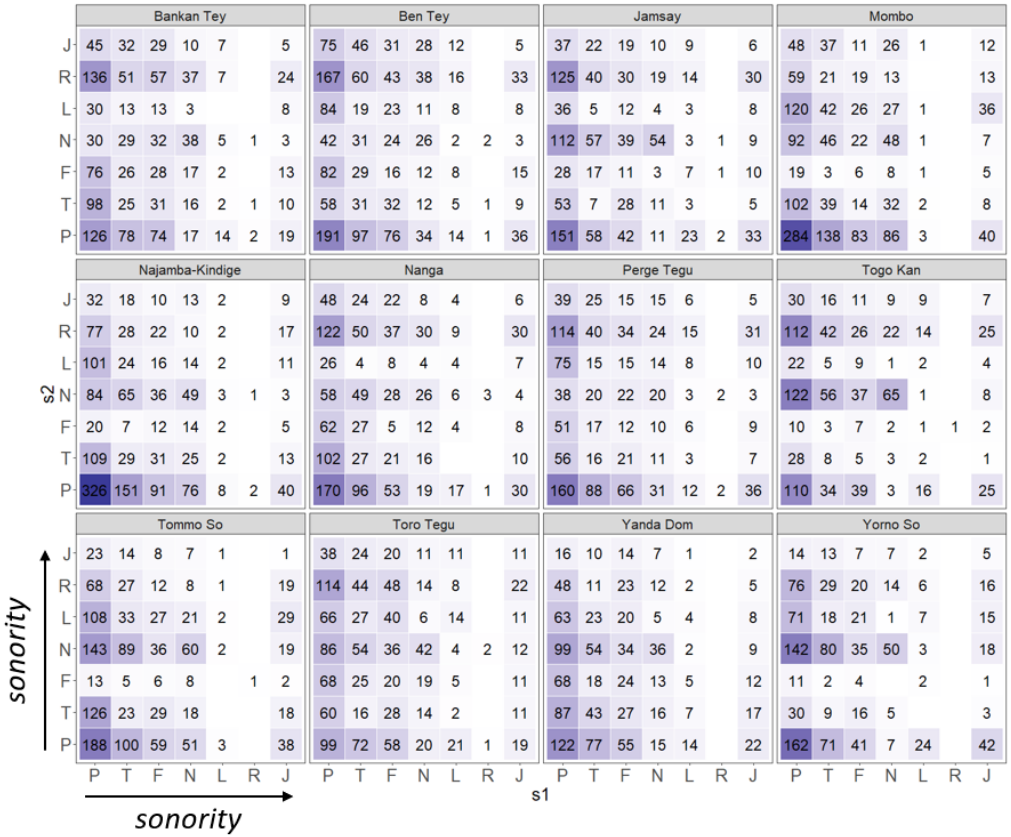

Fig. 2. $\sigma 1$ v. $\sigma 2$ onsets: 2σ words

Disyllabic words

- Gradient effect in $\sigma 1$; R practically non-existent
- General preference for stops in $\sigma 1$ onset
- P…P is most frequent profile in several languages…
- …but note frequency of N & R $\sigma 2$ onsets, esp. in Togo Kan and Yorno So

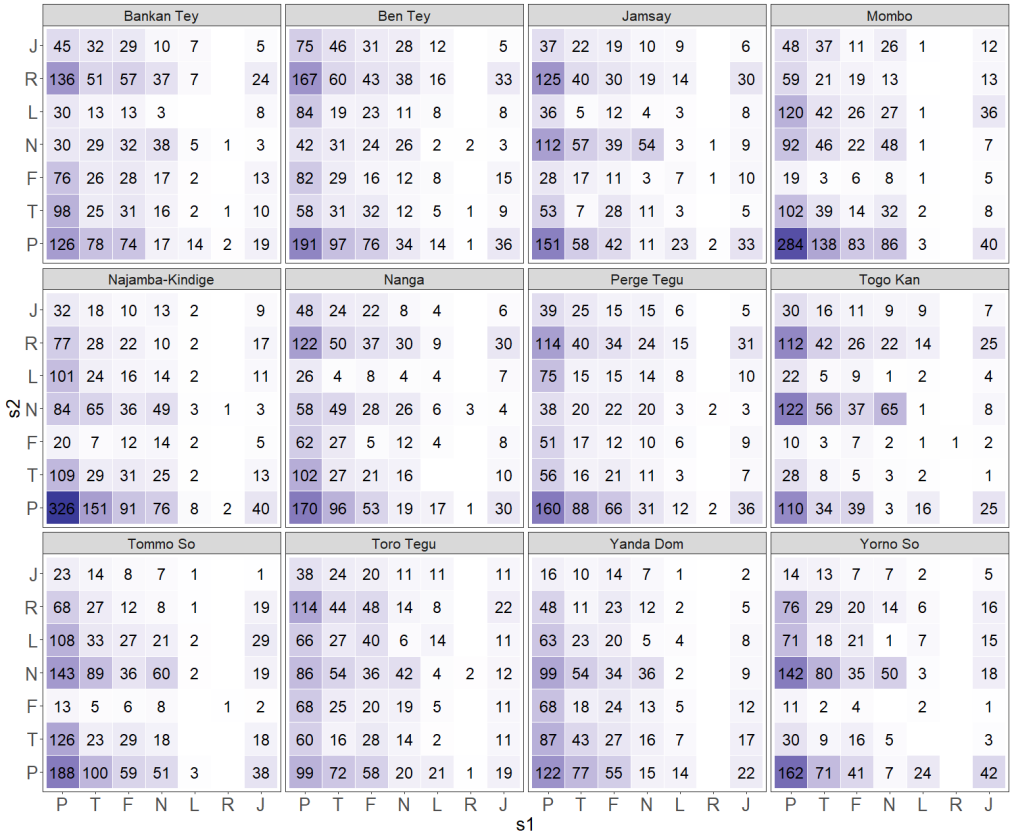

Fig. 3. $\sigma 1$ v. $\sigma 2$ onsets: 2σ words

$\sigma 1 - \sigma 2$, trisyllabic words

- Stops again preferred in $\sigma 1$ onsets.
- Higher-sonority $\sigma 3$ onsets appear less prevalent.

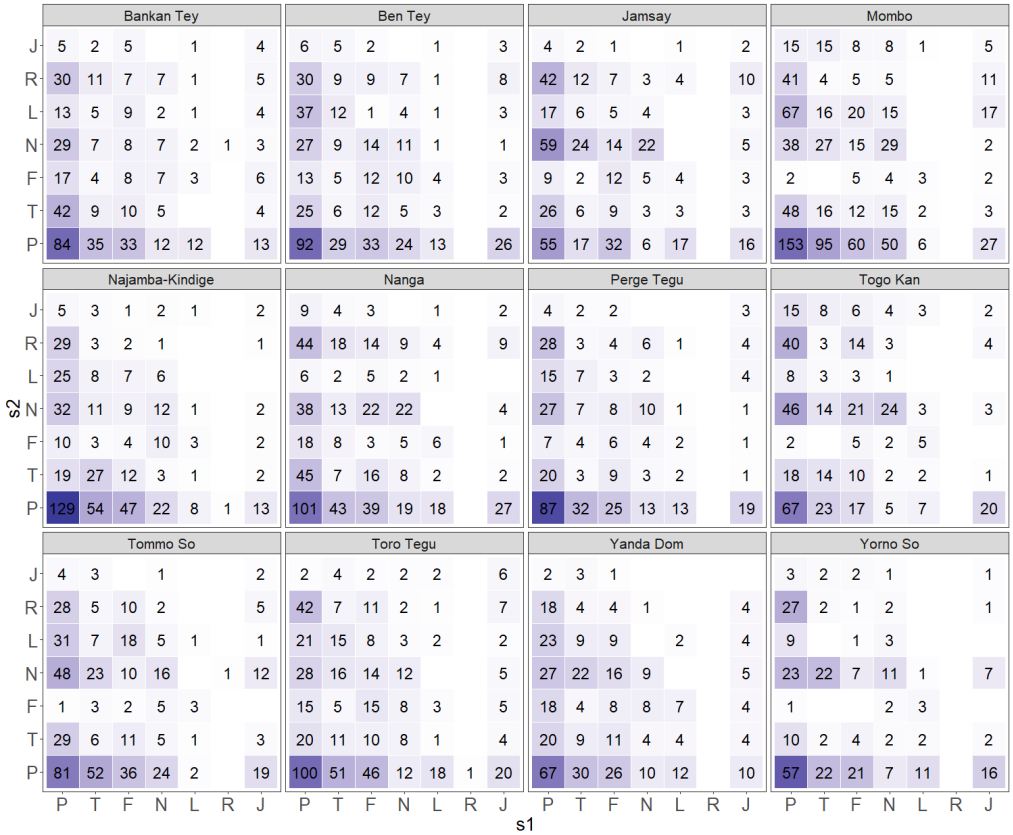

Fig. 4. $\sigma 1$ v. $\sigma 2$ onsets: 3σ words

$\sigma 1 - \sigma 2$, 4-syllable words

- Marked decrease in total number of forms
- The relationship is less clear but seems to be either P…R or split between P…P and P…R

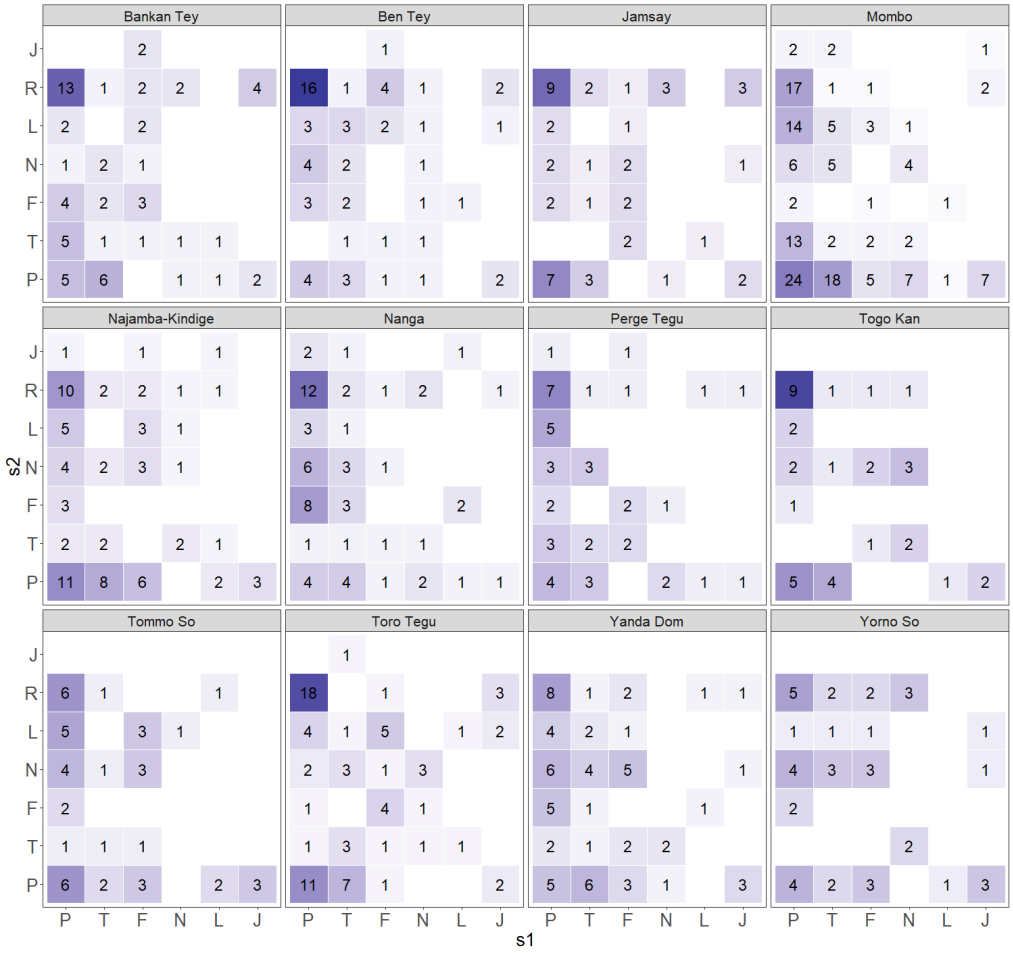

Fig. 5. $\sigma 1$ v. $\sigma 2$ onsets: 4σ words

$\sigma 2 - \sigma 3$, trisyllabic words

- Preference for $\sigma 3$ onset R in most languages
- Greater prevalence of P and/or L in $\sigma 3$ onsets of Mombo, Najamba-Kindige, Tommo So and Yanda Dom
- Most languages disprefer liquid co-occurrences, with exception of: L. . . R in Ben Tey, Perge Tegu and Toro Tegu; L. . . L in Mombo and Tomo So; R. . . R in Toro Tegu

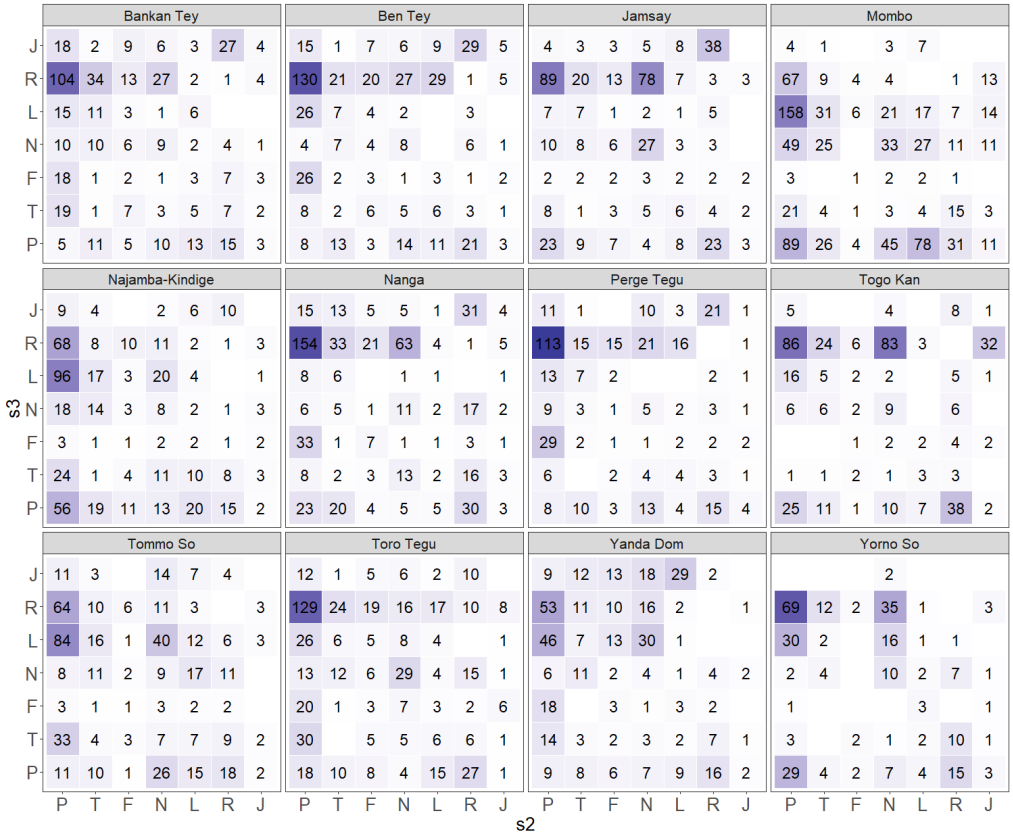

Fig. 6. $\sigma 2$ v. $\sigma 3$ onsets: 3σ words

$\sigma 2 - \sigma 3$, 4-syllable words

- Much more varied, likely due to relative rarity
- Unclear trends in $\sigma 2$ onset; preference for P in $\sigma 3$ in almost all languages
- Clear R. . . P profile in Bankan Tey, Ben Tey and Toro Tegu (these are the same that showed clear $\sigma 1$-$\sigma 2$ P. . . R preference)
- Mombo again exceptional (preference for $\sigma 3$ L)

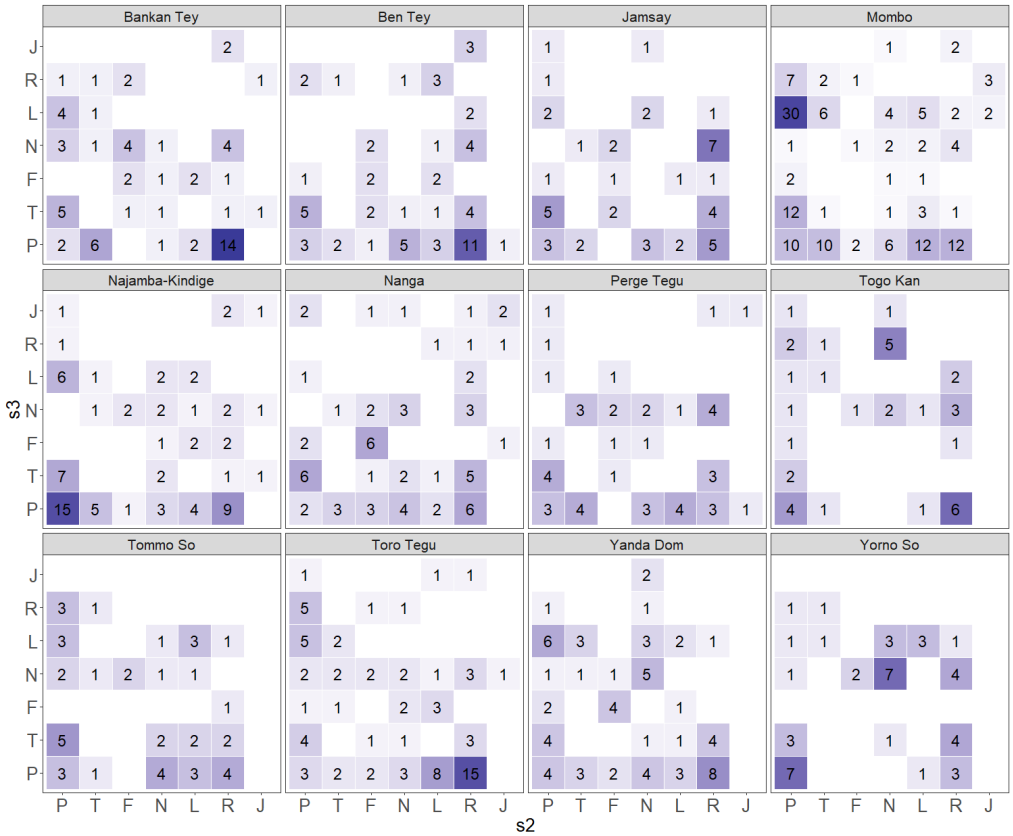

Fig. 7. $\sigma 2$ v. $\sigma 3$ onsets: 4σ words

$\sigma 3 - \sigma 4$, 4-syllable words

- General reemergence of P…R pattern, with some exceptional cases
- 4-syllable trends in monomorphs should generally be taken with a grain of salt

Fig. 8. $\sigma 3$ v. $\sigma 4$ onsets: 4σ words

Discussion

Summary

- Low-sonority onsets are preferred in **first position** in all word sizes.
- Trisyllabic words also show a greater affinity for low-sonority onsets in **second syllables**, whereas many languages prefer R in this position in even-parity forms (other languages preferring other types of high-sonority onsets).
- **Third-syllable** onsets are generally R (or higher sonority) in trisyllabic forms and stops in 4-syllable forms.
- **Fourth-syllable** onsets return to a preference for higher sonority (esp. R).

So what?

- **In sum**, a preference for higher sonority onsets (esp. R) is arguably predictable with respect to the right edge in many languages.
- This should not be taken as suggestive that relevant constraints or parameters *actively* shape onset profiles, especially in simple roots.
- However, through generalization over the lexicon, emergent constraints could motivate the alternations that we do see in morphologically complex forms, going back to the processes cited in the introduction.

Implications for prosody

- Dow et al. (2017) explain /r…l/ ⟶ [l…r] in Ben Tey as "fortition" of weak positions in superfeet (cf. Fig. 9)
- I now see this analysis as suggesting these three syllables are strong (foot and superfoot head), weak and semi-weak (foot head but superfoot non-head) respectively.
- Essentially, the lower sonority [l] would be derived in a weaker position.

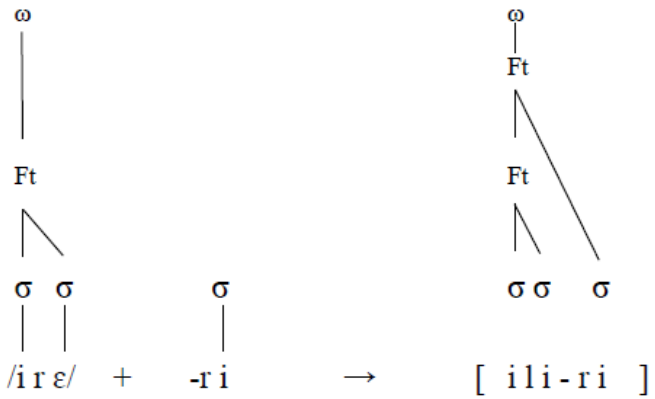

Fig. 9. 'Flip-frop' as $\sigma 2$ fortition

- These facts and the current results make more sense to me now if right-edge syllables are prosodically *weak* and penultimate syllables are *strong*—thus, formation of trochees from right-edge.

Is it Dogon-specific?

- Maybe. The same methodology, applied to French using Lexique (Gimenes et al., 2020), showed a very neat gradient effect within monomorphs.
- We should look at languages with distinctive stress to see if we get similar gradience across the board or positional effects.

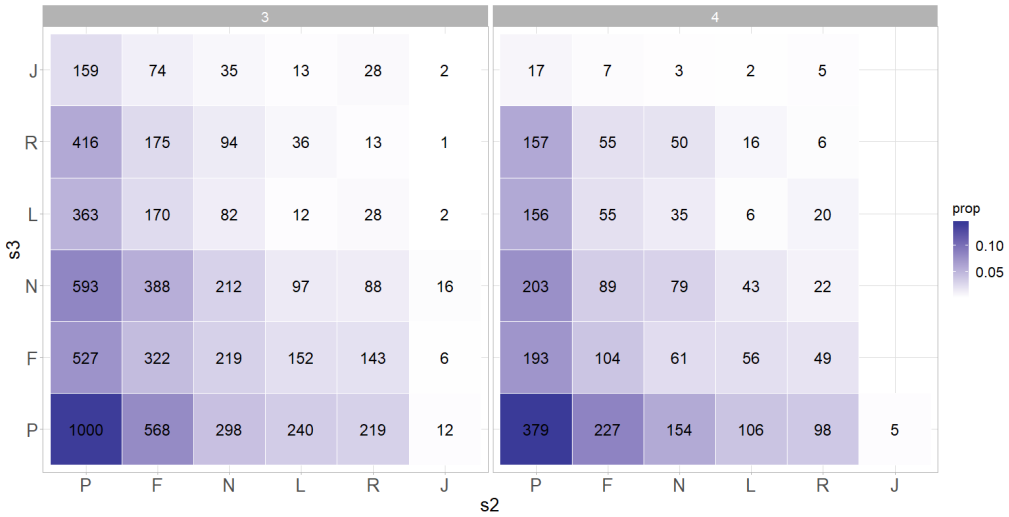

Fig. 10. $\sigma 2$ v. $\sigma 3$ French onsets: 3 & 4σ words

Future work/Open questions

- "Brute force" identification of monomorphs. There could be unencoded or fossilized suffixes.
- Some describe <r> as a flap. Is the prevalence of rhotic onsets just a product of a flapping process? I'm not inclined to think so—I don't see glaring gaps in P…T where P…R is prevalent.
- Dogon languages generally allow only high-sonority codas, and we frequently get word-final <r> as a result of V ~ ∅. If word-final epenthesis is secretly driving up these right-edge high-sonority consonants, we should expect them to pattern more frequently with epenthetic vowel qualities (whether "default" vowels or via vowel copy).

Works Cited I

De Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology.* Cambridge University Press.

Dow, Michael, Christopher Green, and Ryan Hendrickson. 2017. Elucidating Dogon prosodic structure: The case of liquid 'flip-frops' in Beni (Dogon). Annual Conference on African Linguistics.

Gimenes, Manuel, Cyril Perret, and Boris New. 2020. Lexique-infra: grapheme-phoneme, phoneme-grapheme regularity, consistency, and other sublexical statistics for 137,717 polysyllabic french words. *Behavior Research Methods* : 1–9.

Gordon, Matthew. 2005. A perceptually-driven account of onset-sensitive stress. *Natural Language & Linguistic Theory* 23(3): 595–653.

Harris, John. 1997. Licensing inheritance: an integrated theory of neutralisation. *Phonology* 14(3): 315–370.

Heath, Jeffrey. 2008. *A grammar of Jamsay*, vol. 45. Walter de Gruyter.

Heath, Jeffrey. 2015a. A grammar of Ben Tey (Dogon of Beni). Unpublished manuscript.

Works Cited II

Heath, Jeffrey. 2015b. A grammar of Toro Tegu (Dogon): Tabi Mountain dialect. Unpublished manuscript.

Heath, Jeffrey. 2016. A grammar of Nanga: Dogon language family, Mali. Unpublished manuscript.

McPherson, Laura. 2013. *A grammar of Tommo So*, vol. 62. Walter de Gruyter.

Moran, Steven, Robert Forkel, and Jeffrey Heath, eds. 2016. *Dogon and Bangime Linguistics*. Jena: Max Planck Institute for the Science of Human History. URL `https://dogonlanguages.org/`.

Rousset, Isabelle. 2004. *Structures syllabiques et lexicales des langues du monde données, typologies, tendances universelles et contraintes substantielles*. Ph.D. thesis, Université Stendhal-Grenoble III.

Ryan, Kevin M. 2019. Prosodic end-weight reflects phrasal stress. *Natural Language & Linguistic Theory* 37(1): 315–356.

Smith, Jennifer L. 2004. *Phonological augmentation in prominent positions*. Routledge.