On the rarity of postnasalized stops

Juliet Stanton, NYU (stanton@nyu.edu)

Workshop on Sequencing in Phonology, University of Tromsø
November 23, 2023

- From the workshop description:

  "Nasalization or aspiration of stops is realized as preceding or following the closure depending on position. I.e., stops are prenasalized in syllable onsets and postnasalized in syllable codas, while the reverse is observed for aspiration. Thus, inside segments there are no precedence relations among features."

- **My goals today:**

  1. To show you that the above, as stated, is not true – as far as nasalization of stops is concerned.
  2. To make progress towards an explanation as to why contrastive postnasalized stops are far less frequent than contrastive prenasalized stops.

- I'll start by walking through the typology of partially nasal stops, to paint the empirical picture.

1 Partially nasal stops

1.1 Prenasalized stops

- In many languages, nasal-stop sequences have a phonological status distinct from other cluster types.

- The term *prenasalized stop* is often used to describe nasal-stop sequences that pattern like segments, and are in addition contrastive with plain nasals and stops (1, 2).

(1) Sample stop inventory (Lua, Niger-Congo, Boyeldieu 1985)

	Labial	Alveolar	Palatal	"Posterior"
Voiceless	p	t	c	k
Voiced	b	d	ɟ	g
Glottal	ɓ	ɗ	ʄ	ʔ
Prenasalized	mb	nd	nɟ	ŋg
Nasal	m	n	ɲ	

(2) Minimal pairs for Lua prenasalized stops (Boyeldieu 1985:42,44; gloss translations mine)
  a. [bīrī] 'fish fillets (or fish nets)'
     [mbīrī] 'pout'
  b. [mɨ́rā] 'star'
     [mbɨ́rā] 'impatient'

- What do I mean when I say *pattern like segments*?
  - What we call prenasalized stops are often subject to the same phonotactic constraints as segments.
  - In Lua, prenasalized stops can occur word-initially (2), but other consonant sequences do not occur. Potential initial clusters introduced through borrowings are broken up by epenthesis (3).

(3) Initial cluster repair in Lua (Boyeldieu 1985:182)

a. [bòrōòs] 'brush' (from French *brosse*)

b. [kérlé] 'key' (from French *clef*)

- Another criterion often used to determine whether or not a nasal-stop sequence is a segment is whether or not it is *separable* (see Riehl 2008).
  - Western Dani (Austronesian, Barclay 2008, (4)) allows voiced stops following nasals, but voiced stops don't independently exist. Nasal-stop sequences thus aren't separable, and are analyzed as segments.

(4) Western Dani consonant inventory

	Bilabial	Alveolar	Velar	Labiovelar
Prenasalized	mb	nd	ŋg	ŋgʷ
Voiceless	p	t	k	kʷ
Nasal	m	n		
Approximant	w	ɾ		

  - This analysis is backed up by phonotactics. While nasal-stop sequences can occur word-initially (as in words like [mbeɾe] 'two'), other word-initial sequences are forbidden (*[kɾi]).

- Nasal-stop sequences can arise as allophones of other segments. These sequences behave like segments.
  - Nasal-stop sequences often arise as allophones of voiced stops.
    - ⋗ In Qaqet (Baining, Tabain et al. 2022), initial voiced stops are variably prenasalized.
    - ⋗ When present, prenasalization lasts around 60-70 ms. (out of a total duration of around 100ms).
  - Nasal-stop sequences arise as allophones of plain nasals when preceding an oral vowel. This alternation (examples in (5)) is part of a larger phenomenon called *environmental shielding*.

(5) Environmental shielding in Chiriguano (Tupian, Dietrich 1986, gloss translations mine)

a. /m/ → [mb] [amboáku] 'I warm up' (Dietrich 1986:61)

b. /m/ → [m] [ãmõtãta] 'I toughen up' (Dietrich 1986:61)

- **To summarize**: nasal-stop sequences that act like segments are frequent in the world's languages. Sometimes they are allophones of stops or nasals, and sometimes they are contrastive with these segment types.

1.2 Postnasalized stops

- As is true for nasal-stop sequences, stop-nasal sequences can occur as allophones of other segments.
  - Another type of environmental shielding occurs when plain nasals following oral vowels are realized as stop-nasal sequences (examples in (6)).

(6) Environmental shielding in Nadëb (Nadahup, Barbosa 2005, gloss translations mine)

a. /n/ → [dn] [ʃədn] 'hair' (Barbosa 2005:42)

b. /n/ → [n] [napiɟ̚] 'sieve' (Barbosa 2005:45)

  - Stop-nasal sequences also arise as allophones of nasals in a number of Australian languages; this phenomenon is usually referred to as *prestopping* (examples in (7)) and can also affect laterals.

(7) Prestopping in Arabana-Wangkangurru (Pama-Nyungan, Hercus 1972, with comparisons)

Arabana-Wangkangurru	*Bāgundji*	Gloss
d̪idna	dina	'foot'
gudna	gun	'excrement'

- *Postnasalized stops*, or stop-nasal sequences that are contrastive with nasals and other segment types, are attested in a few languages. One well-documented case comes from Alyawarra (Arandic, Yallop 1977).

– Alyawarra's consonant inventory contains oral stops, postnasalized stops, and nasals (8). Minimal (or near-minimal) triplets illustrate these contrasts ((9), conversions to IPA by me).

(8) Consonant inventory of Alyawarra (Yallop 1977:12)

	Labial	Dental	Alveolar	Retroflex	Palatal	Velar
stops	p	t̪	t	ʈ	c	k
postnasals	pm	t̪n̪	tn	ʈɳ	cɟ	kŋ
nasals	m	n̪	n	ɳ	ɲ	ŋ
laterals		l̪	l	ɭ	ʎ	
other continuants	w		r	ɽ	j	(ɰ)

(9) Demonstration of plosive contrasts (Yallop 1977:13)
a. [apiɽa] 'river gum' (eucalyptus species)
   [apmiɽa] 'camp, place'
   [amiɽa] 'woomera, spear-thrower'
b. [atuma] 'hit, kill'
   [atnima] 'digging stick'
   [anima] 'sit, stay'

– One argument for segmental status is that postnasalized stops can occur in clusters (examples in (10)).

(10) Clusters including postnasalized stops in Alyawarra (Yallop 1977:17)
a. [irpma] 'scar'
b. [atntirima] 'run'
c. [arkŋa] 'blood'
d. [uʈɳtima] 'dance'

– Triconsonantal clusters are otherwise unattested unless the third member is [w], as in [lkw] (Yallop 1977:43). (Postnasalized stops can also occur in [w]-final clusters, as in [rkŋw]).

• While it may be true that the arguments for segmenthood of stop-nasal sequences in Alyawarra aren't rock solid, they are just as good as most of the arguments offered for prenasalized stops in other languages.

Returning to sequencing, the theme of this workshop:

• In Alyawarra, postnasalized stops contrast with nasal-stop sequences (11).

(11) Minimal pairs for partially nasal sequences (all from Yallop's 1977 lexicon)
a. [ampira] 'around, round and round, in a circle'
   [apmira] 'place, camp'
b. [intima] 'smell (give off a smell)'
   [itnima] 'stand, be standing, be (situated)'
c. [akŋima] 'carry, take, bring'
   [aŋkima] 'talk, speak, make a noise '

• Yallop treats stop-nasal sequences as segments and nasal-stop sequences as clusters, a decision that is echoed for closely related Central Arrernte by Breen (2001) (see also Round 2023).

• A clearer suggestion that ordering matters *within the segment* comes from Nemi (New Caledonia; Rivierre 1975; Ozanne-Rivierre 1995a).

– Nemi's consonant inventory includes oral, nasal, prenasalized, and postnasalized stops (Table 1).

– There aren't any explicit arguments that the partially nasal sequences are segments, but other segment sequences aren't transcribed or discussed (in Rivierre 1975 or Ozanne-Rivierre 1979, 1995a).

Table 1: Partial consonant inventory of Nemi (adapted from Ozanne-Rivierre 1995a)

		Labialized	Labial	Coronal	Palatal	Velar
Stops	Voiceless	p^{w}	p	t	c	k
	Aspirated	p^{hw}	p^{h}	t^{h}		k^{h}
	Postnasalized	pm^{w}	pm	tn	cɲ	kŋ
	Prenasalized	mb^{w}	mb	nd	ɲɟ	ŋg
Nasals	Voiceless	m̥^{w}	m̥	n̥	ɲ̊	ŋ̊
	Voiced	m^{w}	m	n	ɲ	ŋ

- **To summarize**: just as is the case for nasal-stop sequences, stop-nasal sequences can occur as allophones of plain nasals, or as segments that contrast with plain nasals and stops.

1.3 A frequency asymmetry

- While both prenasalized and postnasalized stops exist, and can contrast with each other, there is a large asymmetry in their frequency (Maddieson & Ladefoged 1993).
  - In UPSID (a balanced sample of 454 languages), 55 languages (12%) have prenasalized stops.
  - By contrast, 2 languages (.004%) have postnasalized stops.
- Most proposals that restrict the predicted typology of contour segments can't explain this asymmetry.
  - **Aperture Theory** (Steriade 1993) seeks to limit the composition and size of contour segments.
    - ⩾ Explains: why only plosives (stops and affricates) can be partially nasalized. Plosives have two featural docking sites (the closure and release); continuants do not.
    - ⩾ Predicts: contour stops can only have two phases (counterexample: Karitiâna [bmb], Storto 1999).
  - **Q theory** (Inkelas & Shih 2017, Shih & Inkelas 2018) seeks to limit the size of contour segments.
    - ⩾ Explains: why there can be tripartite (and longer) contour segments (e.g. Karitiâna [bmb]).
    - ⩾ Predicts: contour segments can contain any combination of subparts (counterexample: *nl, etc.).
  - Neither can explain why postnasalized stops are more frequent than prenasalized stops.
- Today I'll put some meat on a proposal due to Herbert (1978), which focuses on this asymmetry. In brief:
  - Contour segments are created through a process called unification, whereby the pieces of a cluster are fused into a segment (see also Herbert 1986).
  - Prediction: the types of contour segments that we see should mirror the types of clusters that we see.
    - ⩾ Nasal-stop sequences should be frequent, as prenasalized stops are frequent.
    - ⩾ Stop-nasal sequences should be infrequent, as postnasalized stops are infrequent.

<u>Roadmap</u>

- ***First:*** a concrete mechanism by which clusters are unified into segments (Gouskova & Stanton 2021).
  - In the initial state, a learner assumes that all consonant sequences are clusters.
  - Consonant clusters with a specific statistical distribution, where the members are more likely to occur together than they are apart, are unified into complex segments.

- ***Second:*** stop-nasal sequences are much less likely to be unified than nasal-stop sequences.
  - Cross-linguistically, stop-nasal sequences are far less frequent than nasal-stop sequences.
  - Stop-nasal sequences also tend to be more diverse in their composition than nasal-stop sequences. In cases where this is true, unification is less likely.

2 A learning-based approach to complex segmenthood

- Making sense of the typology of complex segments involves understanding what they are and how they arise. Gouskova & Stanton (2021) have proposals in these areas, in the form of a computational learner.
- Our learner borrows ideas from Herbert (1986) and Riehl (2008), but implements them in novel ways.
  - **From Herbert (1986)**: the idea that complex segments result from the unification of clusters. What's novel: for us, unification is based on statistical distributions.
  - **From Riehl (2008)**: the idea that a primary diagnostic for complex segments is separability. What's novel: for us, separability is probabalistic and gradient.
- It's easiest to demonstrate how the learner works with a case study. For reasons of time, I'll just walk through the structure of the learner, but the case study (Boumaa Fijian) is provided for you in gray boxes.

2.1 The initial state

- We assume that the learner has access to a lexicon. Each word is composed only of simplex segments.
  - **Simplex segments** are those that can be characterized using unique and non-conflicting feature specifications for place and manner.
    - ⩾ [m] is labial and nasal throughout its articulation.
    - ⩾ [t] is oral throughout its articulation and involves a single constriction at the alveolar ridge.
  - **Complex segments** are those where specifying either place or manner of articulation can only be done with multiple specifications.
    - ⩾ Labiovelars (e.g. [kp]): constrictions at multiple places of articulation.
    - ⩾ Prenasalized stops (e.g. [mb]): sequenced nasal and oral portions.
- We assume also that the learner has access to each segment's distinctive features.
  - Standard assumption in much work on computational phonology (e.g. Hayes & Wilson 2008).
  - Only important here in that it allows the learner to distinguish between vowels and consonants.

Figure 1: Learning data in the initial state: all consonant sequences are clusters

a m b a n d o n i
ð a n r a
t a ð i n d a r u
n d a u l i β i
s e a ŋ g a ŋ g a
e n d ʒ i t a

2.2 Some background on Boumaa Fijian

- Boumaa Fijian's consonant inventory, proposed by Dixon (1988:13), is below.

(12) Segment inventory of Boumaa Fijian (Dixon 1988:13)

	Labial	Dental	Postalveolar	Velar	Glottal
Nasal	m	n		ŋ	
Stop	p	t		k	ʔ
Prenasalized stop	mb	nd		ŋg	
Prenasalized trill		nr			
Fricative	f, β	s, ð			
Affricate			tʃ, ndʒ		
Liquid		l, r			
Glide			j	w	

- Why are Fijian affricates and prenasalized segments are treated as complex segments?
  - There are no other licit consonant sequences. Treating stop-fricative and nasal-stop sequences as segments allows for a simplified description of the phonotactics: all syllables are (C)V.
  - Not all complex segments are separable (for example, there are no independent voiced stops).
- Our Fijian lexicon was from the An Crúbadán project (`crubadan.org`) and totaled 17,600 words.

2.3 The inseparability measure

- The learner's first act is to calculate an inseparability measure for each cluster (C_1C_2).
- To calculate this measure, the learner analyzes the lexicon for the following:
  - Probability of each C_1C_2 (number of C_1C_2 / total number of biconsonantal clusters).
  - Probability of each C_x (number of C_x / total number of consonants, in any environment).
- Inseparability takes into account the probability that both C_1 and C_2 occur in C_1C_2.
  - **Forward** inseparability is the probability of C_1 being in C_1C_2.

(13) $Insep_{forward}\ (C_1C_2)$: $\frac{Prob(C_1C_2)}{Prob(C_1)}$

  - **Backward** inseparability is the probability of C_2 being in C_1C_2.

(14) $Insep_{backward}\ (C_1C_2)$: $\frac{Prob(C_1C_2)}{Prob(C_2)}$

  - **Bidirectional** inseparability (or just *inseparability*) is the product of (13) and (14).

(15) $Insep_{bidir}\ (C_1C_2) = Insep_{forward} * Insep_{backward}$

  - A bidirectional measure is necessary because, in some cases, one consonant has a restricted distribution but the other does not. The right measure should reflect this.
- Bidirectional inseparability will be high when...
  - The numerators are high: consonants do not combine relatively freely; the language has few clusters.
  - The denominators are low: the segments are not very frequent overall.

- To calculate inseparability measures for Fijian C_1C_2 sequences, we need the relative frequency of individual consonant phones (Table 2), as well as the relative frequency of each C_1C_2 (Table 3).

Table 2: Individual phone frequencies for Fijian

p	1137	b	2328	m	6339	f	394	β	8834	w	1202
t	8372	d	2512	n	7653	s	4871	ð	2467	r	4947
k	9824	ɡ	1026	ŋ	2281	ʃ	1985	ʒ	320	l	6092
ʔ	14									j	1001
											total: 73,599

Table 3: Inseparability measures and CC counts for Fijian

sequence	CC frequencies
ŋ ɡ	1026
m b	2328
n d	2512
t ʃ	1985
d ʒ	320
n r	708
	total: 8,879

- Calculation for [ŋ ɡ], the most inseparable of the sequences in (3):

(16) $$\frac{1026/8879}{2281/73599} * \frac{1026/8879}{1026/73599} = \frac{.1156}{.031} * \frac{.1156}{.0139} \approx 3.729 * 8.317 \approx 30.91$$

  - Forward inseparability = 3.729, backward inseparability = 8.317.
  - Backward inseparability is higher because [ɡ] does not occur outside [ŋ ɡ].

- Calculation for [n r], the most separable sequence:

(17) $$\frac{708/8879}{7653/73599} * \frac{708/8879}{4947/73599} = \frac{.0797}{.104} * \frac{.0797}{.0672} \approx 0.77 * 1.19 \approx 0.91$$

  - Forward inseparability = 0.77, backward inseparability = 1.19.
  - Backward inseparability higher because [r] is less frequent than [n] is outside [n r].

- In words: [n r] has lower inseparability because its pieces occur fairly independently. [n ɡ] has higher inseparability because its pieces don't. The other clusters fall somewhere in the middle.

2.4 The unification procedure

- After calculating inseparability measures, the learner converts all eligible sequences into complex segments. To be eligible for unification, a sequence must satisfy two criteria.

1. Cluster inseparability must be ≥ 1. (We found that a setting of 1 consistently leads to interpretable results, but this setting could be treated as a parameter of the model.)
2. Cluster frequency must be significantly different than zero. (This helps the learner deal with residue.)

- After assembling a list of eligible sequences, the learner makes several sequenced changes to its inventory and representation of the learning data.
  1. Learner modifies its inventory; adds new complex segments and their features.
  2. Learner modifies its lexicon, by iteratively unifying eligible sequences, from most to least inseparable.
  3. Learner checks that each segment included in its inventory is still present in the lexicon, and removes any absent segments from the inventory.
     – For example, in Fijian, [g] is not attested outside of [ŋg].
     – Once the learner unifies [ŋ g], [g] is no longer attested, so it's removed from the inventory.

2.5 Iteration

- Following unification, the learner iterates the procedure until it finds no more eligible clusters.
- Iteration is necessary for several reasons (one obvious one: some complex segments have more than two parts). But iteration isn't relevant to the questions at hand here, so we'll skip it.

2.6 Typological predictions

- This proposal helps us understand several aspects of the the typology of complex segments, because it makes a statement about what complex segments are: at their hearts, they're just a type of cluster.
- One property of complex segments that has received a lot of attention is that they're short. We can derive this without needing to make arbitrary representational limits (as in Steriade 1993, Inkelas & Shih 2017).
  – Complex segments are often short. Most contain two subsegments, sometimes three, and rarely four.
  – Clusters are also often short, even in languages where longer clusters are in principle licit.
    ⋗ Not many languages allow clusters with four or more members (counts in (18) calculated from Gordon 2016's 98-language sample; see his p. 91).

(18) Maximum cluster size across 98 languages

Maximum cluster size	1	2	3	4	5	6
Number of languages	34	25	16	6	8	8

    ⋗ Even in languages that allow long clusters, they are uncommon. Russian, for example, allows clusters up to five members. But clusters with more than two members are infrequent.

(19) Intervocalic clusters in Russian (from Tikhonov 1996)

Sequence	Raw count	Percentage
VCV	182,397	62.9%
VCCV	93,883	32.3%
VCCCV	11,604	4.00%
VCCCCV	1,915	0.66%
VCCCCCV	31	0.01%

    ⋗ A study by Rousset (2004), looking at the frequency of syllable types in 16 languages, makes the same point: there is likely an inverse correlation between cluster length and frequency of attestation.
  – If long clusters are rare, then it follows that long complex segments are also rare.
- Another popular property of complex segments is that they're pretty limited in terms of their composition.
  – Prenasalized stops are common, but prenasalized laterals ([nl]) and trills ([nr]) are rare.[1]
  – We can use our proposal to make a prediction. If we're right that complex segments are unified clusters, then [nr] and [nl] should be rare clusters as well.

[1] If they exist. Prenasalized trills are reported in a few languages (including Fijian), but I don't know of any prenasalized laterals.

3 The distribution of postnasalized stops

- A quick recap of the proposal and prediction detailed above:
  1. Complex segments are really just consonant clusters with a particular statistical distribution.
  2. The typologies of complex segments and clusters should, in certain ways, mirror each other.
- Postnasalized stops are less common than prenasalized stops, so we might expect stop-nasal clusters to be differently distributed than nasal-stop clusters. Next, I'll show that this expectation is borne out.

3.1 Methodology

- To get a sense of the relative distribution of stop-nasal and nasal-stop clusters, I analyzed 24 of the lexica used by Gouskova & Stanton (2021). The corresponding languages and their families are listed below.[2]

Table 4: Languages consulted

Family	**Language**
Athabaskan	Navajo
Austronesian	Fijian (Boumaa) Sundanese
Aymaran	Aymara (Bolivian)
Indo-European	English (American) French (European) German Greek (Modern) Italian Polish Russian Spanish (Mexican)
Mongolic	Mongolian (Khalka)
Niger-Congo	Ngbaka Shona Tswana Wolof
Nilo-Saharan	Mbay
Pama-Nyungan	Wargamay Yindjibarndi
Quechuan	Quechua (Bolivian)
Semitic	Hebrew (Modern Standard)
Turkic	Turkish
Uralic	Hungarian

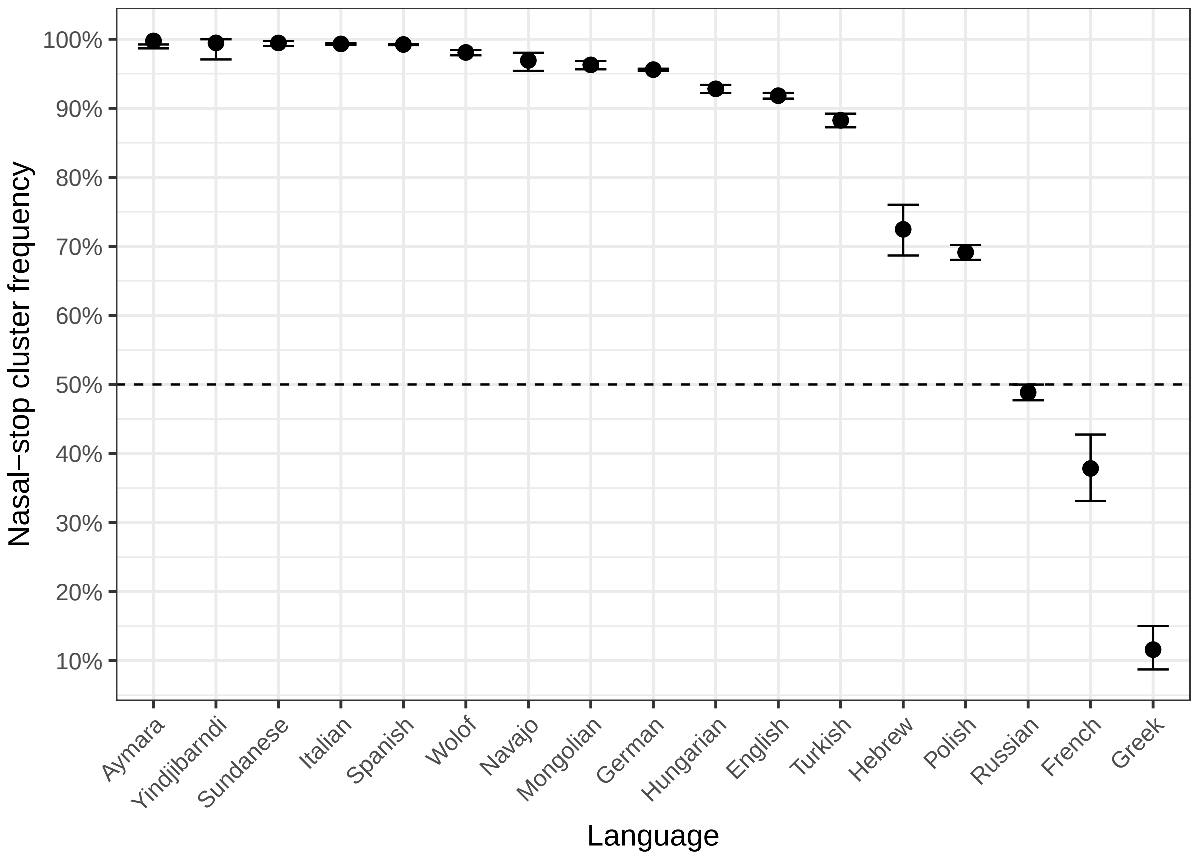
- For each language, I identified all homorganic and heterorganic nasal-stop and stop-nasal clusters composed of labials and alveolars (20). (Why labials and alveolars? Because all 24 languages have them.)

(20) Target clusters

First consonant	Homorganic	Heterorganic
Nasal	[mb], [mp], [nd], [nt]	[md], [mt], [nb], [np]
Stop	[bm], [pm], [dn], [tn]	[dm], [tm], [bn], [pn]

[2] For the lexica and sources, see `http://compseg.lingexp.org`.

Figure 2: Frequency of nasal-stop clusters across languages



3.2 Frequency of nasal-stop and stop-nasal sequences

- All 24 of the languages allow nasal-stop clusters. 17 of these languages also allow stop-nasal sequences.
  - The 7 languages that don't: Fijian, Mbay, Ngbaka, Quechua, Shona, Tswana, Wargamay.
  - Note that about half of the languages permitting stop-nasal clusters are Indo-European. Because of this, I don't want to make claims regarding the relative frequency of languages that allow these clusters.
- For the majority of the languages that allow both cluster types, nasal-stop sequences are far more frequent.
  - If consonants combined freely, we would expect 50% of the clusters to be nasal-stop clusters and the other 50% to be stop-nasal clusters.
  - But this isn't what happens. This is obvious from Figure 2: while Russian sits at around 50%, the other languages are pretty far from it. (Error bars represent the 95% binomial confidence interval.)
- What's going on with French and and Greek?
  - In French, there are diachronic reasons why nasal-stop sequences are uncommon (see e.g. Azra 1995).
    - ⋗ Sometime between the 10th and 16th centuries, vowels were nasalized preceding a coda nasal.
    - ⋗ These coda nasals later deleted. Putting the two steps together: VNC $\rightarrow$ $\tilde{\text{V}}$C.
    - ⋗ Many of the nasal-stop sequences in the lexicon are in loanwords (Macintosh, Challenger, boyfriend, windsurf, etc.). Nasal-stop sequences aren't illicit; language change just obliterated the native ones.
  - In Greek, the apparent lack of nasal-stop sequences reflects a transcription choice.
    - ⋗ Nasal-stop sequences vary allophonically with voiced stops. This lexicon transcribes these as voiced stops, but it's not totally clear to me that this is the right choice.

> Given that there are few if any languages that contrast nasal-stop segments and clusters (Riehl 2008:64-69), the lack of nasal-stop clusters in Greek is expected.
> The lexicon has 50 items with nasal-stop clusters, all [mps]. I suspect the [p]s are intrusive stops.

- **The overall picture** is one in which (all else being equal) nasal-stop sequences are more frequent than stop-nasal sequences, within and across languages.

3.3 Composition of nasal-stop and stop-nasal sequences

- It's well known that nasal-stop clusters are largely homorganic, often as a result of place assimilation (Mohanan 1982, Itô 1986, Padgett 1995, among many others).
  - This makes nasal-stop clusters less separable than they might otherwise be, because it introduces predictability as to what nasals precede what stops and vice versa.
  - if [p] or [b] (for example) is preceded by a nasal, that nasal will always be [m].
- To the best of my knowledge, no work has systematically investigated whether or not a similar homorganicity bias is apparent for stop-nasal clusters.
- Determining this involves more than just comparing the numbers of homorganic and heterorganic clusters.
  - When comparing the frequencies of clusters, necessary to control for the frequencies of their members.
  - If [m] is far more frequent than [n], and [d] far more frequent than [b], then it wouldn't be particularly surprising if heterorganic [md] and [dm] are the most frequent.
- One way to assess a cluster's frequency, while controlling for the frequencies of its members, is to calculate the cluster's Observed/Expected ratio (Pierrehumbert 1992, cf. Wilson & Obdeyn 2009).
  - The *observed* number for a cluster is the number of times that cluster appears in the lexicon.
  - The *expected* number for a cluster is the number of times we might have expected that cluster to appear in the lexicon, given the frequencies of its individual members.
    > To calculate the expected number for a cluster C_1C_2, we need the overall probability of C_1 and C_2.
    > The expected number: probability of C_1 * probability of C_2 * total number of relevant clusters.
    > Here, the relevant clusters are limited to nasal-stop and stop-nasal clusters.
  - If the O/E is over 1, the cluster is overattested relative to naive expectation. If the O/E is under 1, the cluster is underattested relative to naive expectation.
- I calculated the O/E for each of the clusters in (20).

Results (you can look at the plots while I narrate!)

- Figures 3 & 5: The nasal-stop clusters pattern as expected: homorganic clusters are overattested, while heterorganic clusters are underattested. This pattern also holds in each individual language.
- Figures 4 & 6: The stop-nasal clusters pattern differently; there is no homorganicity preference. But a look at the individual languages suggests a lot of variation.
- While Figure 4 suggests that homorganic stop-nasal sequences are underattested (relative to expectation and to the heterorganic sequences), I'm hesitant to make any sort of claim about this.
  - In some languages, there are idiosyncratic limits on clusters created at word boundaries.
    > Many of the English words with heterorganic stop-nasal clusters contain the suffix *-ment* (e.g.. *appointment*, *contentment*). Many others have the prefix *ad-* (e.g., *admiral*, *administer*).

⋗ ***-ment*** is a Latinate suffix that attaches primarily to Latinate roots (Marchand 1969:331).

⋗ Latinate non-suffixal final consonants mostly with [s], [n], [r], and [l]. The labials ([p], [b], [f], [m], [w]) are either rare or unattested in this position (Cser 2012)

⋗ ***ad-*** is a Latinate prefix (many English words with this prefix were borrowed from Old or later French). In Latin, *ad-* was often realized as *an-* before [n] (e.g., *annuate*). (OED)

⋗ Native English derivatives don't follow this pattern (e.g., *adnominal*), but they're less common.

– The sample isn't geographically or typologically balanced. There are claims that other languages allow only homorganic stop-nasal sequences (see e.g., Round 2023:116 on Pama-Nyungan).

- **Conservatively:** nasal-stop clusters are largely homorganic, but stop-nasal clusters aren't.

Figure 3: Overall O/E measures for homorganic and heterorganic nasal-stop sequences

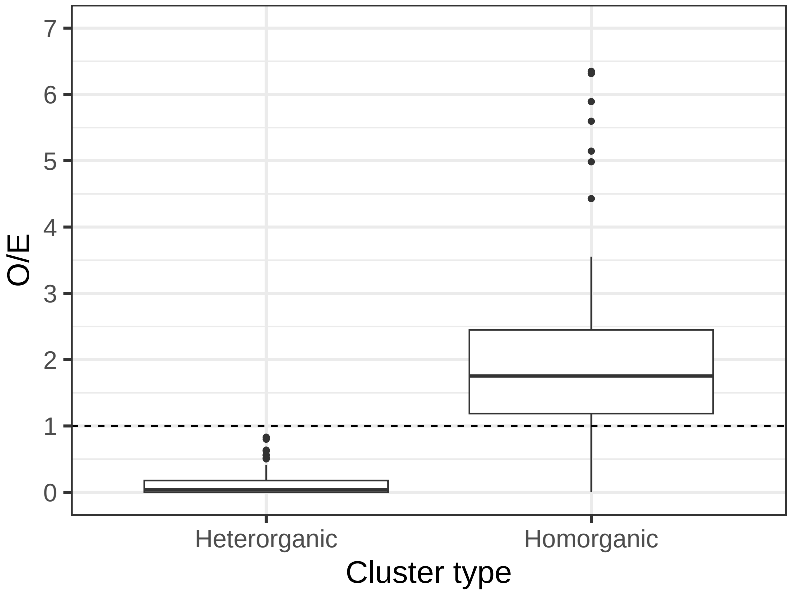


Figure 4: Overall O/E measures for homorganic and heterorganic stop-nasal sequences

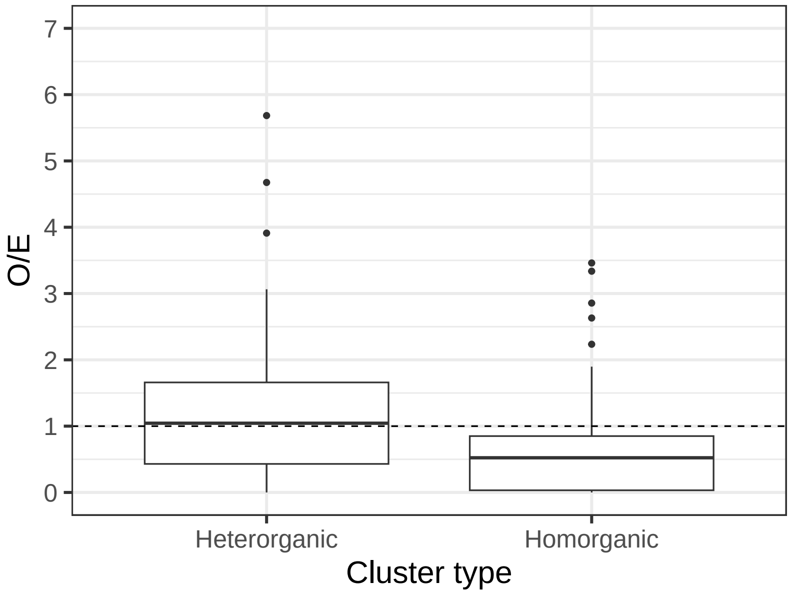


Figure 5: O/E for homorganic and heterorganic nasal-stop clusters, by language

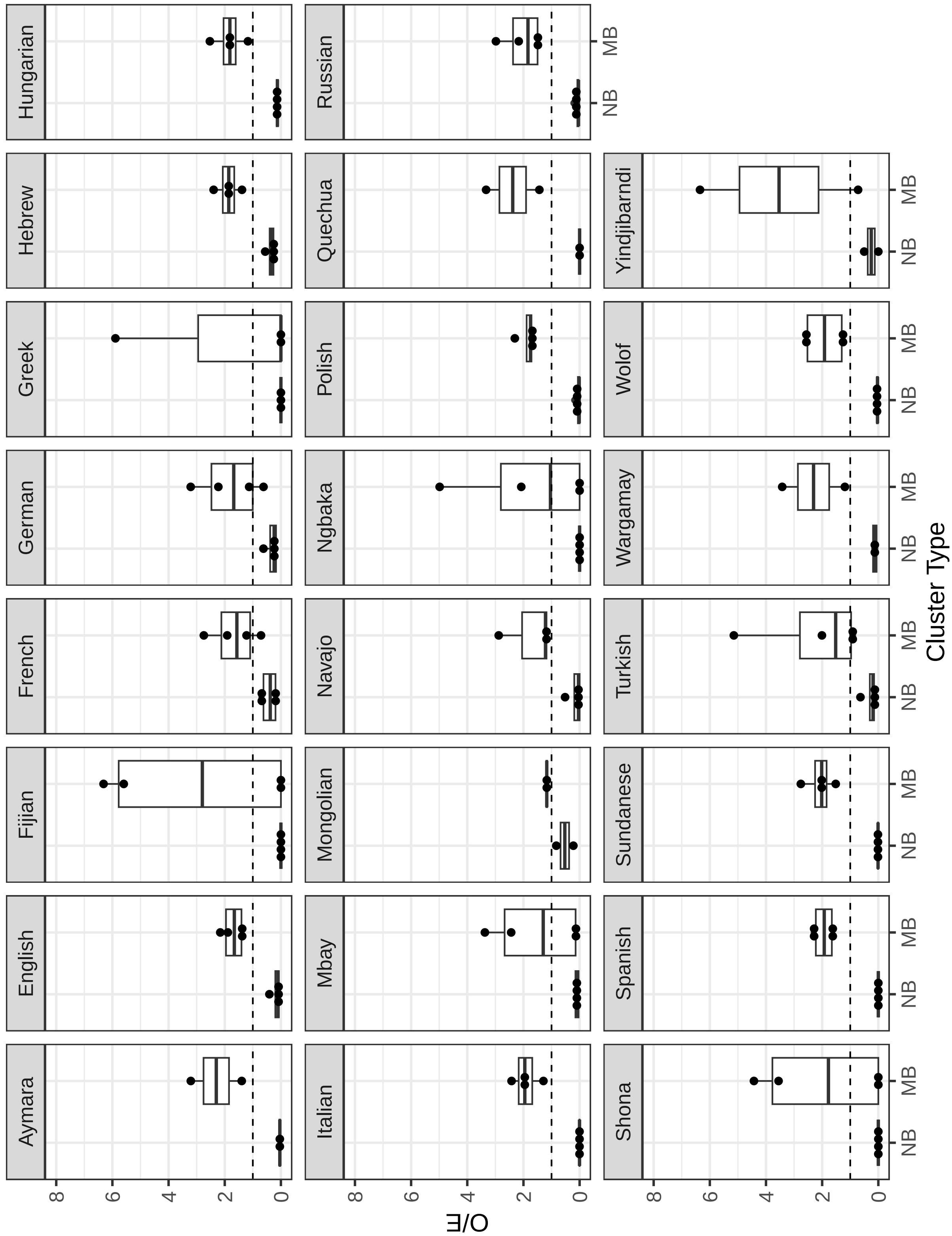
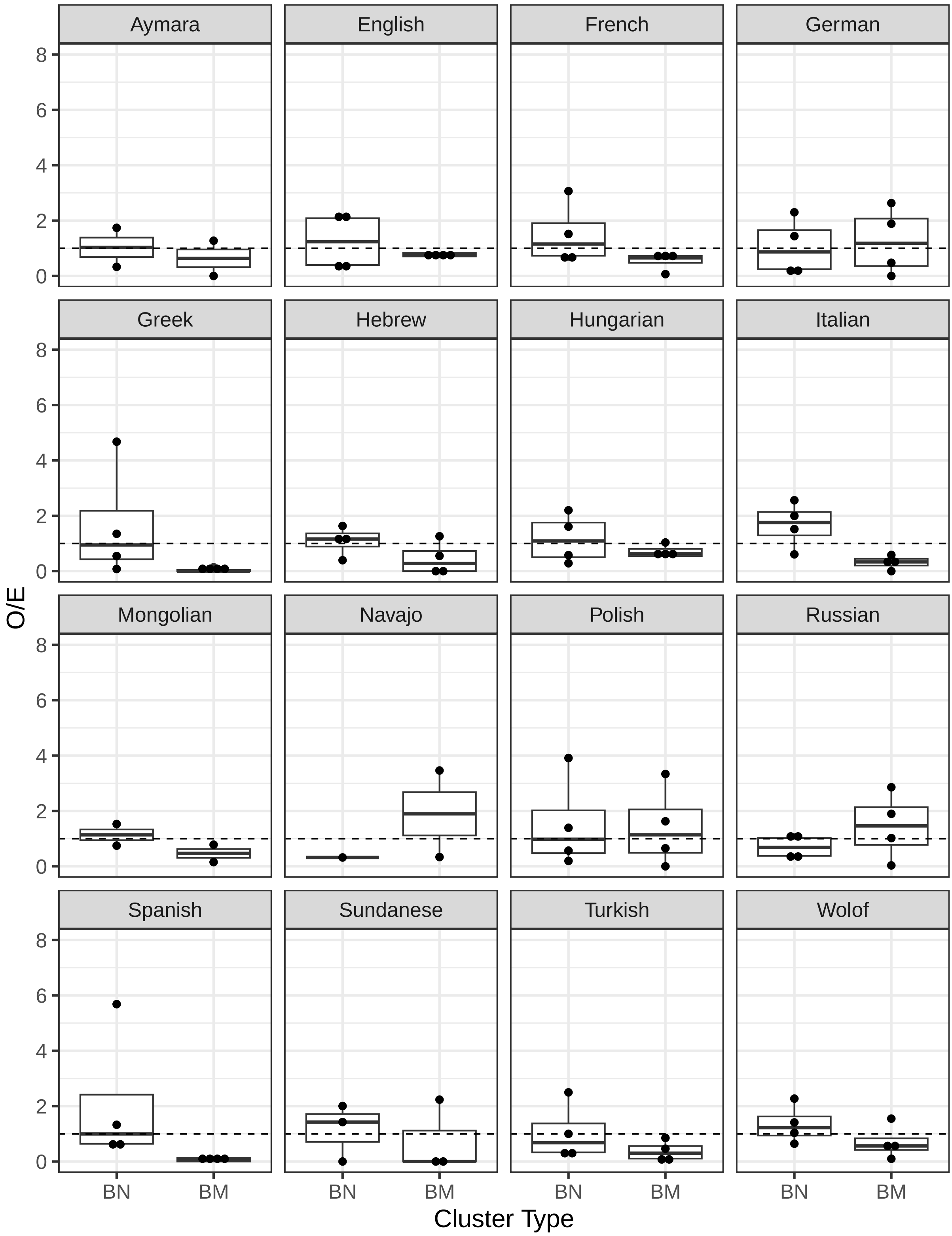


Figure 6: O/E for homorganic and heterorganic stop-nasal clusters, by language



- **Side note**: this asymmetry between nasal-stop and stop-nasal clusters is not surprising.
  - Most analyses of nasal place assimilation hinge on the nasal preceding the stop.
  - There are various ideas about why order matters. Two examples: coda nasals cannot be independently specified for place (Itô 1986), a nasal's place is difficult to perceive preconsonantally (Jun 1995).
  - These proposals don't predict that stop-nasal clusters should behave the same way.

4 Discussion

- Main point of Section 3: stop-nasal sequences exhibit properties that likely cause their overall inseparability to be lower than that of nasal-stop sequences. This is why postnasalized stops are rarer.
  - Stop-nasal sequences are nonexistent in some languages and infrequent in others.
  - Stop-nasal sequences are more diverse in their composition than are nasal-stop sequences: while nasal-stop sequences are largely homorganic, stop-nasal sequences usually don't have such a restriction.
  - Why these facts matter: both would lead to small numerators in the inseparability measures (see (15) for a reminder of how this calculation works), and thus low inseparability measures overall.
- But remember: some languages have contrastive postnasalized stops. How does this happen?
  - In Nemi, there are several sources of words in the lexicon that have stop-nasal sequences (Haudricourt 1962, Rivierre 1975, Ozanne-Rivierre 1995b).
    - ⋗ In Nemi a few and related languages, pretonic vowels were deleted.

(21) Pretonic vowel deletion created stop-nasal sequences

Proto-Oceanic	*Nemi*	*Gloss*	*Source*
*tina	[tne-]	'mother'	Haudricourt 1962:460
*qaɲam	[kŋam]	'pandanus mat'	Ozanne-Rivierre 1995b:54
*tama	[tna-]	'father'	Haudricourt 1962:460

    - ⋗ Verbs can take a nasal infix, which occurs directly following the first consonant.

(22) Stop-nasal sequences created through infixation (Rivierre 1975:354)
   a. [pm^{w}ai] 'to do'
   b. [kŋunduk] 'to drink'
   c. [tnoon] 'to run'

    - ⋗ Nouns can take a prepositional nasal prefix, which assimilates to the first consonant.

(23) Stop-nasal sequences created through prefixation (Rivierre 1975:355)
   a. [kŋa] 'dans la maison' (from ŋa, 'maison')
   b. [pmaac] 'au sec, au terre' (from maac, 'récif')

    - ⋗ A number of onomatopoeias contain stop-nasal sequences.

(24) Examples of Nemi onomatopoeias (Ozanne-Rivierre 1995b:55)
   a. [cɲie] 'to sneeze'
   b. [kŋu] 'to stink'
   c. [kŋaok] 'heron'

    - ⋗ In all cases, place assimilation creates homorganic stop-nasal clusters. This homorganicity requirement, in addition to the low overall number of clusters, would raise these clusters' inseparability.
  - In Alyawarra and other Arandic languages, a likely precursor is a process of prestopping (as in (7)).
    - ⋗ Prestopped nasals and laterals are widespread in Australian languages, and occur as post-tonic allophonic variants of nasals and laterals in many of the languages related to Alyawarra.
    - ⋗ It's likely that the contrastive postnasals of Alyawarra and others arose from this allophonic process.
    - ⋗ Evidence for the plausibility of this change: prestopping appears to be becoming contrastive for other languages of the region (see Harvey et al. 2019 on Arabana-Wangkangurru).
- **My belief**: attempts to explain the typology of complex segments by appealing to constraints on intra-segment sequencing, or the grammar (representational or computational) more generally, aren't fruitful.
- Rather, we can make sense of the typology by adopting the view that complex segments are clusters with a specific distribution, and then taking a good look at language-general constraints on segment sequencing.

References

Azra, Jean-Luc. 1995. Emergence and Evolution of French Nasal Vowels: Reconsidering data through the interplay of production and perception. In *Historical linguistics 1995 : selected papers from the 12th International Conference on Historical Linguistics*, 1–22. Amsterdam: John Benjamins.

Barbosa, Jefferson Fernando. 2005. *Descrição fonológica da língua nadêb*. Brasília: Universidade de Brasilia MA thesis.

Barclay, Peter. 2008. *A grammar of Western Dani*. München: Lincom Europa.

Boyeldieu, Pascal. 1985. *La langue lua ('niellim')*. Cambridge: Cambridge University Press.

Breen, Gavin. 2001. The wonders of Arandic phonology. In Jane Simpson, David Nash, Mary Laughren, Peter Austin & Barry Alpher (eds.), *Forty years on: Ken Hale and Australian languages*, 45–69. Pacific Linguistics.

Cser, András. 2012. The role of sonority in the phonology of Latin. In Steve Parker (ed.), *The Sonority Controversy*, 39–63. Berlin: De Gruyter Mouton.

Dietrich, Wolf. 1986. *El Idioma chiriguano: Gramática, textos, vocabulario*. Madrid: Instituto de Cooperación Iberoamericana.

Dixon, R. M. W. 1988. *A grammar of Boumaa Fijian*. Chicago: University of Chicago Press.

Gordon, Matthew K. 2016. *Phonological Typology*. Oxford: Oxford University Press.

Gouskova, Maria & Juliet Stanton. 2021. Learning complex segments. *Language* 97. 151–193.

Harvey, Mark, Nay San, Margaret Carew, Sydney Strangeways, Jane Simpson & Clara Stockigt. 2019. Prestopping in Arabana. *Australian Journal of Linguistics* 39(4). 419–462.

Haudricourt, André G. 1962. Les consonnes postnasalisées en Nouvelle Calédonie. In *Proceedings of the Ninth International Congress of Linguists*, 460–461.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440.

Herbert, Robert K. 1978. The Distributional Frequency of Half-Nasal Consonants. In V. P. De Guzman & R. K. Herbert (eds.), *Calgary Working Papers in Linguistics* 4, 27–44. LOGOS.

Herbert, Robert K. 1986. *Language Universals, Markedness Theory, and Natural Phonetic Processes*. Berlin: Mouton de Gruyter.

Hercus, Louise Anna. 1972. The Pre-Stopped Nasal and Lateral Consonants of Arabana-Waŋgaŋuru. *Anthropological Linguistics* 14. 293–305.

Inkelas, Sharon & Stephanie S. Shih. 2017. Looking into Segments. In Karen Jesney, Charlie O'Hara, Caitlin Smith & Rachel Walker (eds.), *Proceedings of the 2016 Annual Meeting on Phonology*, 1–18. Linguistic Society of America.

Itô, Junko. 1986. *Syllable theory in prosodic phonology*. Amherst, MA: University of Massachusetts Amherst dissertation.

Jun, Jongho. 1995. *A Constraint Based Analysis of Place Assimilation Typology*. Los Angeles: University of California, Los Angeles dissertation.

Maddieson, Ian & Peter Ladefoged. 1993. Partially nasal consonants. In Marie Huffman & Rena Krakow (eds.), *Nasals, nasalization, and the velum*, 251–301. San Diego, CA: Academic Press.

Marchand, Hans. 1969. *The categories and types of present-day English word-formation*. München: Verlag C. H. Beck.

Mohanan, K. P. 1982. *Lexical phonology*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Ozanne-Rivierre, Françoise. 1979. *Textes nemi (Nouvelle-Caledonie)*. Paris: SELAF.

Ozanne-Rivierre, Françoise. 1995a. Nemi. In Darrell T. Tryon (ed.), *Comparative Austronesian Dictionary: An Introduction to Austronesian Studies*, chap. 64, 849–854. De Gruyter Mouton.

Ozanne-Rivierre, Françoise. 1995b. Structural Changes in the Languages of Northern New Caledonia. *Oceanic Linguistics* 34(1). 44–72.

Padgett, Jaye. 1995. Partial class behavior and nasal place assimilation. In *Arizona Phonology Conference 5: Proceedings of South Western Optimality Theory Workshop*, 145–183.

Pierrehumbert, Janet. 1992. Dissimilarity in the Arabic Verbal Roots. In *Proceedings of the 23rd Meeting of the Northeastern Linguistic Society*, Amherst, MA: GLSA.

Riehl, Anastasia K. 2008. *The phonology and phonetics of nasal obstruent sequences*. Ithaca, NY: Cornell University dissertation.

Rivierre, Françoise. 1975. Phonologie du némi (Nouvelle Caledonie) et notes sur les consonnes post-nasalisées. In *Bulletin de la Société de Linguistique de Paris*, 345–356. Librairie C. Klinksieck.

Round, Erich. 2023. Phonotactics. In Claire Bowern (ed.), *The Oxford guide to Australian languages*, 106–119. Oxford: Oxford University Press.

Rousset, Isabelle. 2004. *Structures syllabiques et lexicales des langues du monde*. Grenoble: Université Grenoble dissertation.

Shih, Stephanie S. & Sharon Inkelas. 2018. Autosegmental Aims in Surface Optimizing Phonology. *Linguistic Inquiry* 50(1). 137–196.

Steriade, Donca. 1993. Closure, release, and nasal contours. In Marie Huffman & Rena Krakow (eds.), *Nasals, nasalization, and the velum*, 401–470. San Diego, CA: Academic Press.

Storto, Luciana. 1999. *Aspects of a Karitiana grammar*. Cambridge, MA: MIT dissertation.

Tabain, Marija, Marc Garellek, Birgit Hellwig, Adele Gregory & Richard Beare. 2022. Voicing in Qaqet: Prenasalization and language contact. *Journal of Phonetics* 91(101138).

Tikhonov, Aleksandr Nikolaevich. 1996. Morpho-orthographic dictionary: Russian morphemics. Online: `http://www.speakrus.ru/dict/index.htm#tikhonov`.

Wilson, Colin & Marieke Obdeyn. 2009. Simplifying subsidiary theory: statistical evidence from Arabic, Muna, Shona, and Wargamay. Ms. Johns Hopkins University.

Yallop, Colin. 1977. *Alyawarra: An Aboriginal language of central Australia*. Canberra: Australian Institute of Aboriginal Studies.