Learning Complex Segments*

Maria Gouskova and Juliet Stanton, NYU

Abstract

Sequences such as [mb, kp, ts] pattern as complex segments in some languages but as clusters of simple consonants in others. What evidence is used to learn their language-specific status? We present an implemented computational model that starts with simple consonants, and builds more complex representations by tracking statistical distributions of consonant sequences. This strategy succeeds in a wide range of cases, both in languages that supply clear phonotactic arguments for complex segments and in languages where the evidence is less clear. We then turn to the typological parallels between complex segments and consonant clusters: both tend to be limited in size and composition. We suggest that our approach allows the parallels to be reconciled. Finally, we compare our model with alternatives: learning complex segments from phonotactics and from phonetics.

1 Introduction

1.1 Quechua vs. French

Languages differ in the representational status of sequences such as [tʃ]. In Quechua, there are a number of arguments that [tʃ, tʃ', tʃʰ] pattern as affricates, part of the natural class of stops. First is an argument from inventory structure: Quechua has a three-way laryngeal contrast in stops, [p t k p' t' k' pʰ tʰ kʰ], and the affricates complete the series. The fricatives [ʃ, s, x], on the other hand, do not have laryngeal contrasts—[ʃ'] and [ʃʰ] do not occur except in affricates. The second argument is from syllable phonotactics: Quechua disallows word-initial CC and medial CCC, and bans stops in codas (see 1a). If [tʃ, tʃʰ, tʃ'] were stop-fricative clusters rather than affricate stops, they would be the sole exception—why is [k'at͡ʃa] allowed but not [kat͡sa]? The third argument is from nonlocal restrictions on ejective and aspirated stops. These stops cannot follow other stops within the same phonological word, and affricates are no different (see 1b). If [ʃʰ, ʃ'] were separate segments, their distribution would require a convoluted statement: they can—in fact must be—immediately preceded by [t], as in [sat͡ʃ'a], but they cannot be preceded by other stops at a longer distance, *[pat͡ʃ'a]. Thus, the inventory and phonotactics can be analyzed more elegantly if [tʃ, tʃʰ, tʃ'] are complex segments.

*We received useful feedback from Andries Coetzee, Michael Becker, Lisa Davidson, Gillian Gallagher, Maddie Gilbert, Donca Steriade, and an anonymous reviewer. Thanks to audiences at MIT, AMP 2019, and NYU, as well as Adam Albright for Latin paradigms, Michael Becker for Turkish materials, and Maxim Kisilier for the Modern Greek corpus. This work was supported in part by NSF BCS-1724753 to the first author.

(1) Quechua affricates (data from Gallagher 2016, 2019, Gouskova & Gallagher 2020)

a. Basic phonotactics: no stops in coda; no tautosyllabic clusters

misk'i	'delicious'	*mi͡ksi, *ka͡tsa	but cf.	k'a͡tʃa	'pretty'
tʃ'impu	'boil'	*t͡simpu, *ti͡ntsu	but cf.	aɲt͡ʃa	'a lot of'

b. Laryngeal co-occurrence constraint: no [stop. . .C'] or [stop. . .C^h]

tʃ'uspi	'to fly'	*tʃ'usp'i	*p'uk'i
satʃ'a	'tree'	*patʃ'a	*kap'a
k^{h}utʃi	'pig'	*k^{h}utʃ^{h}i	*k^{h}ut^{h}i

By contrast, European French has only simplex segments (Fougeron & Smith 1993). French allows [tʃ], which may appear in word-initial position in a few loanwords: [tʃad] 'Chad', [tʃetʃɛn] 'Chechen'. But French [tʃ] is not special: [ps] and [ks] can occur word-initially, too ([psikoloʒi] 'psychology', [kseʁɛs] 'sherry'). French phonology does not offer any arguments that [tʃ] is anything but a cluster of a stop followed by a fricative.

If the distinction between complex segments and clusters of simplex segments is a real difference in the mental representations, then these representations must be learned: the French speaker must discover that [tʃ] is two consonants, while the Quechua speaker must discover that [tʃ] is an affricate. We ask how learners figure this out.

1.2 The learning problem

The learning problem can be decomposed into two questions: (i) what language-internal cues do learners use to discover complex segment representations, and (ii) when does this happen?

The issue of language-internal cues is what we consider in this paper. It is a nontrivial problem, since even to analysts, the treatment of complex segments is not always straightforward. Starting as far back as Trubetzkoy 1939 and Martinet's (1939) response, the heuristics have been controversial. Trubetzkoy's criteria set the stage for the subsequent developments in this field. Is the duration of the sequence like that of a cluster, or like that of a segment? Is the sequence heterosyllabic or tautosyllabic? Does the sequence have the same distribution as an uncontroversial singleton segment? Is the language's phonemic inventory more symmetric if the sequence is analyzed as a complex segment? Can the sequence be decomposed into parts that occur independently? We explore the last criterion, which we term INSEPARABILITY (following Riehl 2008). Unlike previous work, we define inseparability as a gradient measure: the likelihood of C_1 and C_2 occurring together as C_1C_2, rather than separately or in clusters with other Cs. Our findings indicate that in a range of languages, inseparability is the key to identifying complex segments. Inseparability succeeds both in languages where other heuristics clearly diagnose complex segments and in languages where the arguments for complex segments are less clear or contradictory.

As a proof of concept, we implement our proposal as a computational learner (§2). Our model assumes that in the early stages of phonological acquisition, learners acknowledge only simplex segments. Learners then decide, based on the rates at which consonants occur alone and in clusters, whether each cluster would be better analyzed as a segment—that is, UNIFIED (following Herbert 1986). Our learner captures the difference between Quechua and French, and it can more generally differentiate complex segments from clusters in ways that mirror the conclusions

of analysts. We discuss a range of cases, including Fijian, Mbay, Ngbaka, Turkish, Hebrew, Latin, English, Russian, Sundanese, Shona, and Greek. These languages have a variety of complex segments: affricates, prenasalized segments, labiovelars, labialized consonants. The learner finds all of them, demonstrating that the result is general.

The issue of the learning timeline is harder to address given the available evidence. We do not know when children acquire complex segment representations, or how these representations interact with the learning of phonotactics, alternations, and morpheme segmentation. Complex segment representations are difficult to study because they are covert; attempts to find phonetic or behavioral evidence of complex segments even in experiments with adult participants have not always produced interpretable results (§6.4). Until better evidence comes to light, we can study the timeline by investigating the data available to learners at different acquisition stages: connected speech (especially child-drected speech), segmented phonological words, a morphemic lexicon. There are relatively few languages for which all three types of data are available and where the analytic status of complex segments is clear. In several cases, we get substantially the same results no matter what type of data we use. But, as we discuss in more detail in §2.7, in the few cases where results diverge significantly between data types, the morphemic lexicon emerges as the likeliest data source.

1.3 Typology, phonotactics, and phonetics

We suggest that our learner, coupled with additional assumptions, can explain certain facts about the typology of complex segments. One typological generalization about complex segments concerns their size: they are often composed of two subparts, sometimes three, and rarely four (Steriade 1993, Shih & Inkelas 2019a). Under our proposal, this limitation follows naturally. If complex segments result from unification, we would only expect long segments to be common if long clusters are, too. In reality, long clusters are rare both cross-linguistically and language-internally (see §5). When the learning data supply evidence of complex segments with more than three parts (as in Zezuru Shona), our learner finds them. Our proposal's ability to derive the size generalization—and its potential extensions to other as-of-yet unexplained generalizations—favors it over other current theories of complex segments, which either do not address these generalizations or explain them through stipulation.

We consider some alternative approaches in §6: learning complex segments from phonotactics and phonetics. In the phonotactic approach, ambiguous sequences such as [ts] are a hidden structure problem. Learners construct phonotactic grammars assuming cluster vs. complex segment representations, checking for improvement in fit. We argue against this approach, because the fit of the resulting grammars turns out to always improve as more complex segments are added to the inventory—regardless of whether these segments make sense for the language. Another alternative is phonetics: [ts] is not ambiguous; the learner can detect acoustic or articulatory differences that signal each sequence's status. We review the research looking for such differences, and show that (a) sometimes the phonetic evidence contradicts the phonotactics, and (b) it is not clear that there are reliable asymmetries in the phonetics of all clusters vs. all complex segments (Maddieson 1983, Arvaniti 2007, a.o.).

1.4 The role of representations

We need to clarify our terms. We assume that a SIMPLEX SEGMENT has unique, nonconflicting specifications for place and manner. An [m] is nasal and labial throughout; a [t] has a constriction at the alveolar ridge. On the other hand, in COMPLEX SEGMENTS, either the constriction or the manner of articulation involves two or more distinct specifications.[1] Prenasalized stops and affricates mix manners: [mb] starts as nasal but ends as oral, [ts] involves stop and fricative constrictions. Doubly articulated stops and secondary articulations mix places and/or manners: [kp] involves labial and velar stop constrictions, and [kw] has a velar stop constriction with lip rounding. The one unifying feature of complex segments is that they are articulatorily complex in a way that simplex segments are not, and are decomposable into separate segments in the same language or other languages.

When discussing clusters of simplex segments, we will simply call them CLUSTERS. Complex segments are often written in a special way to highlight analytical assumptions or phonetic differences in duration: the alveolar affricate can be [t͡s], [ts], or [t^{s}]; nasal-stop segments can be [n͡d], [^{n}d], or [n^{d}]. We eschew these conventions—when we need to highlight the difference between a cluster and a complex segment, we use spaces: [m b] is a cluster, and [mb] a segment.

We make no distinction, terminological or otherwise, between complex and contour segments (cf. Sagey 1986). Sagey defines complex segments as having simultaneous articulations, and contour segments as sequential. But this distinction is not universally accepted. Lombardi (1990) argues that affricates—clearly sequential phonetically—have phonologically unordered fricative and stop portions (see Lin 2011 on affricate debates). Even labiovelars such as [kp]—phonetically simultaneous—can phonologically pattern as contours: when nasals assimilate to labiovelars partially, it is the dorsal and never the labial part of [kp] (Padgett 1995a, Van De Weijer 2011). Moreover, the contour-complex distinction rests on analytic arguments as much as on phonetic reality. In labiovelars such as [kp], the two constrictions are near-simultaneous, but their releases must be staggered to be audible (Maddieson 1990, Ladefoged & Maddieson 1996). Phonologically complex labialized stops [kw, tw, pw] are articulatorily heterogeneous: simultaneous lip rounding and closure are impossible for [pw], but this doesn't necessarily affect its analytical treatment. Sequential constrictions are not even guaranteed for clusters: Browman & Goldstein (1989) show that in the English *perfect memory*, [t] is inaudible because it is completely overlapped with surrounding stops). Crucially, all of these sequences present the same problem for the learner.

We are open to representational differences among complex segments. It is even possible that languages differ in whether, say, affricates have sequential or unordered representations, especially if learners construct the representations from language-internal phonological evidence. But we take the question of formal representation to be logically distinct from the question of how the presence complex segments is detected, and concern ourselves only with the latter issue.

Our major claim is that complex segments represent a learner's decision that certain clusters are better analyzed as segments, due to aspects of their distribution. Complex segments are shortcut representations for sequences that have the distributions of segments. These shortcut representations often result in more elegant phonotactic grammars and contrastive inventories, but those are consequences, not goals.

[1] Segments involving different laryngeal configurations (aspirates, ejectives, implosives) are sometimes treated as complex (Kehrein 2013, Shih & Inkelas 2019a); we have no quarrel with this view, but do not treat them as sequences to be unified in our simulations, to keep the paper to a reasonable length.

2 The learner

This section introduces our computational learner (§2.1) and illustrates its application to Boumaa Fijian (henceforth Fijian). The learner has three components: an INSEPARABILITY MEASURE (§2.2), a UNIFICATION PROCEDURE (§2.3), and ITERATION (§2.4). The algorithm is summarized as pseudocode in §2.5; §2.6 discusses the source of the distributions that allow the learner to be successful, and §2.7 considers the issue of the learning data.

2.1 The initial state

The learner gets two types of information in the initial state. First is a lexicon of LEARNING DATA containing only simplex segments. Figure 1 shows a snippet of the initial state lexicon of Fijian. Our Fijian corpus is from An Crúbadán (http://crubadan.org, 26,000 words from internet texts—17,600 after filtering out English).

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

Figure 1: Learning data in the initial state

Second, the learner gets a FEATURE TABLE of all simplex segments; the natural class defined by [-syllabic]—consonants—is extracted for later calculations. In the initial state, Fijian has the consonants in Table 2. The target consonant inventory is in Table 1 (Dixon 1988:13).[2] Dixon posits six complex consonants: [mb, nd, ŋg, nr, tʃ, ndʒ]. The affricates [tʃ, ndʒ] are usually allophones of /t, nd/ before /i/, but can occur in other contexts in loanwords.

[2]Here and throughout, we convert non-IPA sources into IPA according to the descriptions. Dixon characterizes /k/ and /f/ as marginal. We follow Dixon in writing the velar glide as [w]. Fijian orthography writes prenasalized stops as singletons: [mb] = <b>, [nd] = <d>, ŋg = <q>; the other complex segments are tʃi=<ci>, ndʒ=<j>, nr=<dr>. The remaining non-IPA orthographic correspondences are [ŋ] = <g>, [β] = <v>, ð =<c>, ʔ = <’>, j = <y>. All of the orthography-to-IPA conversion scripts, corpora, simulation results, and code for the learner are available on GitHub; see https://github.com/gouskova/transcribers , https://github.com/gouskova/compsegcode, and http://compseg.lingexp.org.

	labial	dental	postalveolar	velar	glottal
nasal	m	n		ŋ	
voiceless stop	p	t		k	ʔ
prenasalized stop	mb	nd		ŋg	
prenasalized trill			nr		
fricatives	f, β	s, ð			
affricates			tʃ, ndʒ		
liquids			l, r		
semi-vowel			j	w	

Table 1: Fijian consonants: target inventory

Dixon's analysis of Fijian rests on a phonotactic argument. Aside from the complex segments, Fijian has no consonant sequences. Complex segments only occur prevocalically, just as singletons. Under this analysis, Fijian has (C)V syllables.

2.2 The inseparability measure

The first step in the learning procedure is to calculate an inseparability measure for each biconsonantal sequence. Intuitively, the inseparability measure tracks how likely a consonant is to be in a specific CC sequence as opposed to other environments—either as a singleton or in another sequence. To assess inseparability, we calculate the probability of each CC sequence, which is the frequency of CC divided by the total number of all CC sequences. We also calculate the probability of each consonant: the number of times the consonant occurs anywhere, divided by the total number of times all the consonants occur. Bidirectional inseparability, 4, is the product of the probability of C_1 being in the cluster C_1C_2 and C_2 being in the cluster C_1C_2. This calculation is essentially the same as Mutual Information (Cover & Thomas 2012), which also tracks the probability of two things co-occurring as a function of their independent probabilities. Our calculation is specific to consonants (that is, [-syllabic] segments; see §6.2). We discuss why the calculation must be over segments rather than natural classes in §6.1.

(2) $\text{Insep}_{forward}(\text{xy}) = \frac{Prob(xy)}{Prob(x)}$

(3) $\text{Insep}_{backward}(\text{xy}) = \frac{Prob(xy)}{Prob(y)}$

(4) $\text{Insep}_{bidir}(\text{xy}) = \text{Insep}_{forward} * \text{Insep}_{backward}$

Our notion of inseparability borrows its name from Riehl's (2008) inseparability criterion, whereby a sequence must be analyzed as a complex segment if at least one of its subparts is not independently attested (cf. Trubetzkoy's 1939 rule VI). For Riehl, Fijian [mb] must be a complex segment, because /b/ does not independently exist. The English sequence [mb], however, does not have to be a complex segment, because /m/ and /b/ both independently exist. The difference between Riehl's conception of inseparability and ours is that our inseparability is probabilistic and gradient: both English and Fijian [m b] have definable inseparability measures.

This bidirectional inseparability measure (henceforth just inseparability) will be very high for any sequence in a language like Fijian, which has complex segments but no clusters. The

reason is that the range of CC sequences in such a language will be fairly limited compared to a language that freely combines consonants in true clusters. Inseparability will also be high if a part of a sequence only occurs in that sequence, or mostly occurs in that sequence. Inseparability will be greater than 1 for any sequence that is more likely to occur as a sequence than as separate parts; taking the product of the two measures in 2 and 3 ensures that the relative freedom of one consonant can be balanced against the boundedness of another. For example, in Fijian, [m] occurs outside of [mb], but [b] does not; the inseparability of [mb] takes into account the distribution of both [m] and [b].

To illustrate, we show the frequencies of individual Fijian phones in Table 2, and the frequencies and inseparability measures of CC sequences in Table 3. Since the learner always looks at bigrams, it only sees the subparts of [n d ʒ]: [n d] and [d ʒ].

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	g	1026	ŋ	2281	ʃ	1985	ʒ	320	l	6092
ʔ	14									j	1001
										total:	73,599

Table 2: Individual phone frequencies for Fijian (first iteration)

sequence	inseparability	CC frequencies
ŋ g	30.91	1026
m b	25.23	2328
n d	22.55	2512
t ʃ	16.29	1985
d ʒ	8.75	320
n r	0.91	708
	total:	8879

Table 3: Inseparability measures and CC counts for Fijian (first iteration)

The calculations leading to inseparability measures for [ŋ g] (the most inseparable cluster) and [n r] (the least inseparable cluster) are presented in detail below. For [ŋ g] in 5, the first term in the equation is the probability of [ŋ g] (given all the clusters) divided by the probability of [ŋ] (given all the consonants). This is roughly 3.729. The second term is the probaility of [ŋ g] divided by the probability of [g]—roughly 8.317. (The probability of [g] is higher than [ŋ] because [g] does not occur as a singleton, while [ŋ] does.) The product of these two terms is **30.91**.

(5) $\frac{1026/8879}{2281/73599} * \frac{1026/8879}{1026/73599} = \frac{.1156}{.0310} * \frac{.1156}{.0139} \approx 3.729 * 8.317 \approx \mathbf{30.91}$

For [n r], in 6, the first term in the equation is the probability of [n r] (given all the clusters) divided by the probability of [n] (given all the consonants). The second term is the probability of [n r] divided by the probability of [r]. The product of these two terms is **0.91**.

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

There is a clear difference in Table 3 between [n r] and the rest of the clusters. This is because [b], [d], [g], [ʃ], and [ʒ] only occur in CC sequences; clusters containing these segments have high inseparability because the denominator on at least one side of the equation is small. By contrast, both parts of [n r] are independently attested, so the denominator is larger and its inseparability is lower.

2.3 The unification procedure

After the learner has calculated the inseparability of each biconsonantal sequence, it decides which clusters to convert to complex segments and which clusters to leave as is. We call this step UNIFICATION, as a nod to Herbert's (1986) proposal that all segment nasal-stop sequences are underlyingly clusters but are unified over the course of the phonological derivation.[3] To qualify for unification, a sequence must satisfy two requirements:

- *Cluster inseparability must be equal to or greater than 1.* To qualify for unification, a biconsonantal sequence must pass the inseparability threshold of 1. The more frequent the cluster is, and the less frequent its subparts are, the higher its inseparability will be.
- *Cluster frequency must be significantly different than 0.* We do not want the learner to be swayed by residue in the data (loanwords, errors/misparses). To make the learner robust in the face of residue, small numbers must be ignored. We ensure this by adding a check to the learner: if the frequency of a cluster is not significantly different from 0 (using a Fisher's Exact Test at $\alpha = .05$), then it is not a candidate for unification.

We set the inseparability threshold to 1 because this setting consistently leads to interpretable results. It is logically simple: C1 and C2 are unified when they occur more often together than apart. This is comparable to how the O/E statistic is interpreted in Frisch et al. (2004:185), for example. The threshold could, however, be treated as a variable parameter of the model, and our computational implementation allows for this.

Given the Fijian biconsonantal sequences in Table 3, [ŋ g], [m b], [n d], [t ʃ], and [d ʒ] have inseparability measures over 1. For each sequence, the learner checks that its total frequency is significantly different from 0. The Fisher's Exact tests are computed off a contingency table that compares (i) the actual frequency of a cluster and all other clusters, (ii) hypothetical frequencies were that cluster unattested. A sample contingency test for [d ʒ] is in Table 4. The attested and hypothetical distributions are significantly different ($p < .001$), so [d ʒ] satisfies both criteria for unification listed above (as do all other inseparable clusters).

[3] While the spirit of the ideas is similar, our proposal differs from Herbert's in important ways. One major difference is the motivation for unification: Herbert proposes that complex segments are created so as to minimize the number of marked syllable types (see his p. 176, and our §6.3 for further discussion). Herbert also claims that complex segments must meet certain requirements (such as homorganicity, for nasal-stop sequences) to be unified; we make no such restrictions, but §5 discusses ways to derive some of these same generalizations regarding the typology of complex segments.

	[d ʒ]	Other clusters
Attested	320	8,559
Hypothetical	0	8,879

Table 4: Sample Fisher's Exact Test for [d ʒ]

After the learner identifies unifiable sequences, it modifies the segmental inventory and its learning data. First, the learner modifies its feature table by adding the new complex segments.[4] Second, the learner modifies its lexicon, unifying eligible clusters in order of inseparability: in Fijian, the learner replaces [ŋ g] with [ŋg], then [m b] with [mb], and so on for each cluster that passes the unification threshold at this stage. If the learner operates on a lexicon of morphemes rather than morphologically complex words, this allows for the possibility of contrast between [m+b] and [mb] at morpheme boundaries; see §2.7.

Unification of one cluster can bleed unification of another. In Fijian, unification of [n d] (with higher inseparability) bleeds unification of [d ʒ]. This is because all instances of [d ʒ] are part of the trigram [n d ʒ]; the trigram is converted to [nd ʒ] so no instances of [d ʒ] remain. Finally, the learner checks that each segment in the feature table is still present in the lexicon, and removes any absent segments from the feature table. Since [b], [d], [g], [ʃ], and [dʒ] do not occur independently, these segments are removed.

2.4 Iteration

Following unification, the learner computes new frequencies for each segment and cluster, as well as inseparability for each cluster. These values for the second iteration of the Fijian learning simulation are in Tables 5–6.

p	1137	mb	2328	m	4011	f	394	β	8834	w	1202
t	8372	nd	2512	n	5141	s	4871	ð	2467	r	4947
k	9824	ŋg	1026	ŋ	1255	tʃ	1985	ʒ	320	l	6092
ʔ	14									j	1001
										total:	65,748

Table 5: Individual phone frequencies for Fijian (second iteration)

sequence	inseparability	CC frequencies
nd ʒ	521.09	320
n r	80.62	708
		total: 1,028

Table 6: Inseparability measures and CC counts for Fijian (second iteration)

Only two clusters remain, [nd ʒ], [n r], so their inseparability measures are high. The total number of clusters has dropped, so the probability of the remaining clusters increases. The sequence [n r] has a lower inseparability because both [n] and [r] occur as singletons. The frequency

[4]The learner also attempts to add feature representations for the new segments, for compatibility with other statistical learners such as Hayes & Wilson 2008. See http://compseg.lingexp.org/help for specifics.

of both clusters is significantly different from 0, so the learner replaces [nd ʒ] with [ndʒ], then [n r] with [nr], and finally removes [ʒ] from its feature table (as it does not occur separately). The third iteration finds no remaining clusters, and the learner converges on the inventory posited by Dixon (1988).

Iteration is necessary for two reasons, both apparent in the Fijian simulation. First, some complex segments are composed of more than two parts. Fijian has the prenasalized affricate [ndʒ]; Zezuru Shona has prenasalized labialized affricates with 4 parts (§5.3). As our learner only examines bigrams, multiple iterations are necessary to allow it to unify tripartite or longer segments. Second, complex segments sometimes contain phones that appear in more than one sequence. In Fijian, for example, [n] belongs to three different complex segments: [nd], [ndʒ], and [nr]. This means that multiple iterations can be necessary for all of these sequences to qualify for unification, as the inseparability measures of some can be too low on the first pass.

There is no limit to the number of iterations. The learner stops when no more sequences qualify for unification. It is thus capable of finding segments that contain three, four, five, or more subparts. In none of our cases, however, does the learner actually find complex segments longer than four parts (see §5.3 on Shona). We discuss how this length-based restriction on complex segments is derived in §5.

2.5 Summary of the algorithm

The algorithm in pseudocode is in 7. In prose, the learner starts with learning data represented as singleton consonants only. The learner calculates inseparability measures for each cluster. If no clusters exceed 1, the starting versions of the learning data and features are the best; no complex segments are added. If any clusters have inseparability exceeding 1, and their frequency is significantly different from 0, they are sorted from most inseparable to least and rewritten, one at a time, as new complex segments. The learning data are checked to remove any segments that no longer occur in the data as a result of unification, and the feature table is adjusted accordingly. The process is repeated until no remaining clusters qualify for unification.

(7) Complex Segment Learning Algorithm

Input: Learning Data with simple segments, FeatureChart describing the segments

i. Count all CC clusters;

ii. Count all singleton Cs;

iii. Calculate insep for all CCs, sort CCs by insep;

iv. If any insep(C1C2)$\geq$1 and freq(C1C2)>0:

Unify C1C2 as a new C3;

Generate composite features for C3, add C3 to FeaturesChart;

Rewrite LearningData, replacing C1C2 with C3;

For any C in C1C2, check if C in LearningData;

If not, remove it from FeatureChart;

repeat from *i.*

v. Else:

return last version of LearningData and FeatureChart and stop.

2.6 Why these distributions?

As we show in §3–4, our learner finds the same complex segments that linguists posit, in a variety of languages. Probabilistic inseparability is a feature of complex segments in many lexicons; intuitively, when a sequence is a complex segment, it is used in the lexicon often. But why do complex segments have these particular distributions? We think there are several reasons.

First, some complex segments historically derive from singletons, so if the original singletons are frequent, so will their complex descendants. At least some of the prenasalized stops in modern Bantu are reconstructed as voiced stops in Proto-Bantu (Nurse & Philippson 2006:148). Affricates are often the endpoint of stop palatalization, as in Slavic (Jakobson 1929, Townsend & Janda 1996:76ff) and Romance. In §4.1, we simulate the Romance change by converting Classical Latin /k, g, t, d/ to [tʃ, dʒ, ts, dz] before front vowels (as in Vulgar Latin). Our learner easily finds all four affricates.

But single-segment origin does not guarantee that a complex segment will be discoverable—distributions can be over-attenuated by chains of changes. Vulgar Latin affricates became fricatives in Old French (Pope 1934:125ff), and modern European French has no complex segments. Quebecois French re-introduced [ts, dz] as allophones of /t,d/ before [i,y] (Béland & Kolinsky 2005), but our computational learner does not consistently discover these affricates. Between Latin and French, additional changes eliminated many of the older stops, and /t, d/ might have become too rare in the relevant environments.

Sound change does not have to be the source of distributions like those of complex segments. Muyang (Central Chadic) has [mb, nd, ndz, ŋg] alongside [b, d, dz, g] and [m, n]. Prenasalized stops were innovated at some point between proto-Chadic (Newman & Ma 1966) and proto-Central Chadic (Gravina 2014), but it is unclear how they were innovated: voiced stops and nasals existed in both proto-languages, so the prenasalized stops could not have been created

through unconditional prenasalization of voiced stops. It may be that in the proto-Chadic dialect that became proto-Central Chadic, nasals and voiced stops occurred more often together than they did independently. Nasal-voiced stop sequences are frequent both cross-linguistically and within languages, while singleton voiced stops can be infrequent. Under our proposal, these distributional differences would be enough to cause learners to analyze nasal-stop sequences as complex segments.

Another factor that affects distributions is language contact, which can change the lexicon so much that certain sequences acquire the distributions of complex segments. Turkish borrowed [dʒ] from other languages so often that our learner readily identifies it as an affricate (§3.3). By contrast, Russian borrows [dʒ] as heterorganic [d ʐ], never unified in our simulations (§4.2). Thus, borrowing can but does not have to be a source of complex segments.

2.7 On the nature of the learning data

We now consider two broader learnability questions: when do learners acquire complex segment representations, and what data do they use? Our Fijian simulation used a dictionary-like list of phonological words. This type of data is easy for linguists to get and is widely used in statistical computational phonology (Zuraw 2000, Hayes & Wilson 2008, et seq.). But is it a well-motivated choice?

There is no consensus on the nature of the learning data in phonological learnability research. Should learning track type or token frequencies? Type frequencies are implicated in morpho-phonological and phonotactic learning (Bybee 1995, Albright & Hayes 2003).[5] Correspondingly, models such as the UCLA Phonotactic Learner learn from type frequencies in orthographic (≈phonological) word lists (Hayes & Wilson 2008, Hayes & White 2013). But Adriaans and Kager (2010) point out that the word lexicons are an idealization, and are themselves a result of learning—word and morpheme boundaries are not given. Their model locates boundaries by tracking token bigram frequencies in connected speech and learning phonotactics simultaneously. In addition to word lists and connected speech, salient alternatives include learning from morphemes, or morphologically segmented words. A morphemic lexicon is assumed in generative theories (Chomsky & Halle 1968, Halle & Marantz 1993). Morphologically segmented words are the best data for finding certain constraints (Gouskova & Gallagher 2020, Gallagher et al. 2019).

A broad view of phonological learnability must reconcile learning phonotactics, segmentation, and representations such as complex segments, which can present a chicken-and-egg problem. Some phonotactic constraints can only be formulated assuming a certain analysis of the segmental inventory (including complex segments), and with reference to morpheme boundaries. So is one of these problems solved first, or are they solved simultaneously? Perhaps learners revise their segmental inventories after they become morphologically aware, much as has been argued for phonotactic grammars (Gouskova & Becker 2013)?

Answering these questions requires a systematic investigation of different kinds of evidence: roots, morphemes, morphologically complex words, connected child-directed speech. We have only tested varied data in a few languages: Quechua (§3.4), Russian (§4.2), Quebecois French,

[5] Bybee argues from morphology: highly irregular patterns *(go/went)* are token-frequent, but regular ones are type-frequent (found in most verbs).

Hungarian and Navajo (project site). Wherever the results differ, they suggest that the right distributions are in morphemes/roots. Quechua is the clearest: the arguments for complex segments are strong, and the statistical distributions in roots, affixes, and morphologically complex words differ so much that the learner's results differ dramatically depending on the training data. But there are other reasons to think morphemes are the right level of analysis. In some languages, morpheme concatenation creates sequences that are either less frequent morpheme-internally (Russian [ts]) or absent altogether (Mbay has [mb] morpheme-internally but [m b] at boundaries, §3.2). To avoid unifying [m b] at boundaries in languages like Mbay, our learner has to operate on morphemes. This should work both for languages where the phonotactics of morphemes differ from those of complex words, and for languages where the distributions are similar (such as Fijian, Dixon 1988:27–29).

Finally, methodologically, data quality matters. The better curated the corpus, the more likely our learner is to find the complex segments that linguists posit. Unsurprisingly, data prepared by lexicographers and linguists (Ngbaka, Shona, Russian, Turkish, Hebrew) yield cleaner results than online text corpora (Mongolian, Wolof). Whenever possible, we try to examine the qualitative, not just the quantitative results from multiple data sources, especially when the results are negative or the arguments are unclear (as in Latin, §4.1). Sometimes, purported complex segments just miss the threshold for unification ([ts, dz] in Quebecois French, [ts] in German—project site); curiously, in such cases, the arguments for complex segments are not especially clear to begin with.

3 Case studies I: the clear cases

The logic of our argument is to show that inseparability is a feature of complex segments in languages where their status is fairly clear. Thus established, the inseparability metric serves as a learnability-theoretic diagnostic in more ambiguous cases. Ultimately, any claim about covert phonological representations should be supported by experimental evidence. Pending convincing experimental evidence for the psychological reality of complex segments, we consider phonological arguments to be a reasonable starting point. Phonological arguments help us identify the target system that the learner should converge on, and they suggest that there is something at stake in getting the inventory wrong.

We have tested our learner on a large range of languages (currently twenty-five). The cases we discuss comprise two analytic groups. In the first group (§3), the arguments for complex segments are clear (Ngbaka, Mbay, Turkish, Hebrew, Quechua). Unlike Fijian, all these languages allow true clusters (though to different extents), and most of their complex segments are separable. Our learner finds the target inventories for all the languages, though in Quechua, the results depend on the kind of data that the learner is exposed to.

In the second group (§4), complex segments have been posited despite a lack of clear arguments (Latin, Russian, English). Unsurprisingly, our learner finds complex segments in some of these languages but not others. We end this section with Sundanese (§4.4), where analysts posit different complex segments from what our learner finds. Additional cases are discussed in the phonetics and typology sections, as well as on the project site.

3.1 Prenasalized consonants and labiovelars in Ngbaka

Ngbaka (Niger-Congo, Maes 1959, Thomas 1963, Henrix 2015) is described as having prenasalized stops, labialized consonants, and labio-velars (see 8). The consonant inventory of Ngbaka is in Table 7.[6]

	labial	dental	palatal	velar	labiovelar	glottal
stops	p, b, ɓ	t, d, ɗ		k, g	kp, gb	
fricatives	v, vw	s, z				h
nasals	m	n, nw	ɲ	ŋ	ŋm	
prenasalized Cs	mb	nd	nz	ŋg	ŋmgb	
liquids		l, r				
glides	w		j			

Table 7: Consonant inventory of Ngbaka, following Maes 1959

One argument for complex segments is phonotactic. They are both common and allowed wherever simplex segments are allowed, namely in word-initial and intervocalic position. As we show below, there are other clusters, but they are rare and limited. Another argument is that some of the complex segments participate in co-occurrence restrictions (Sagey 1986, Rose & Walker 2004, Danis 2017) in a way that would be difficult to capture if they were clusters. As is sometimes the case with co-occurrence restrictions, there is an identity exemption: nonadjacent [b. . .b] and [g. . .g] are overattested in our corpus (Observed/Expected ratios of 3.91 and 4.33, respectively). But there are restrictions on similar but nonidentical stops: [mb] does not occur before [b] (0 occurrences), and [ŋmgb] does not occur before [gb] (also 0 occurrences). It would be difficult to capture these distributions while treating [gb] and [mb] as clusters of [g] and [b].

(8) Distribution of Ngbaka complex segments (Maes 1959; we translated glosses from French)

a.	kpalɛ	'products of the field'	e.	sakpa	'backpack'
b.	mbata	'large indigenous stool'	f.	bambu	'large waist of mothers after birth'
c.	ŋgabolo	'monkey sp.'	g.	fuŋgu	'wheat soup'
d.	ŋmgbanza	'red ants'	h.	gbaŋmgba	'trap with weights'

Our Ngbaka corpus is a digitized version of Henrix's (2015) dictionary, 5571 words. The vast majority of forms in this dictionary do not contain consonant sequences other than those in Table 7, suggesting a limitation on consonant sequences. (This limitation is not explicitly discussed in any resources on Ngbaka available to us.) Henrix 2015 lists each item with other consonant sequences as an ideophone; three examples are [turtur] 'noise produced by scraping' (p. 541), [mbarmbar] 'covered in big spots' (p. 344), and [harkakaː] 'to be rough, stiff' (p. 206).

The computational learner's task is harder in Ngbaka than Fijian: (i) /ŋmgb/ requires at least two iterations to be unified; (ii) the learner must differentiate the complex segments in Table 7

[6] Maes writes (p. 11) that *'l* and *r* are interchangeable; in some words there is a preference for *l* or *r*; one rarely hears *r* as an initial consonant.' We have included both in the inventory as both are in our source, Henrix 2015. We transcribe the prenasalized labiovelar as /ŋmgb/ following Henrix 2015, Danis 2017 (Maes writes it as ŋb). The Ngbaka we discuss is distinct from Ngbaka Ma'bo (Thomas 1963, Sagey 1986, Rose & Walker 2004, and others). Despite being distinct, the languages have similar segmental inventories and phonotactics; for discussion on the relationship between these languages see Danis 2017:51–53.

from the consonant clusters. This is not necessarily straightforward, as not all complex segments are frequent (/nw/ is attested only 14 times), and all consonant sequences are separable, in the categorical sense.

Our learner ran three iterations on Ngbaka. On iteration 1, it unified the following sequences: [nd, gb, ŋg, kp, nz, mb, ŋm, vw]. This is almost the right result: the learner does not unify [nw] and the prenasalized labiovelar stop on its first pass (though it does find two of its subparts: [ŋm] and [gb]). It does not unify [n w] because this sequence's inseparability is too low; it cannot unify [ŋ m g b] because the learner only considers bigrams. Table 8 presents the calculations for iteration 1; sequences to be unified are above the line. There were many marginal clusters in addition to [r w] that all had an inseparability of ≈0.0, N(C1C2)=1, and Fisher's Exact Test p(C1C2) of 1.0. We left them out of the table to save room.

Sequence	insep	N(C1C2)	N(C1)	N(C2)	p(C1C2)
n d	4.22	484	1115	908	0.0
g b	3.95	924	2124	1855	0.0
ŋ g	3.74	767	1350	2124	0.0
k p	2.52	395	1864	605	0.0
n z	2.38	306	1115	643	0.0
m b	1.81	484	1275	1855	0.0
ŋ m	1.57	385	1350	1275	0.0
v w	1.22	51	130	300	0.0
m g	0.97	380	1275	2124	0.0
n w	0.01	14	1115	300	0.0
r h	0.01	2	127	112	0.5
r k	0.0	5	127	1864	0.062
r t	0.0	3	127	800	0.25
r g	0.0	2	127	2124	0.5
r w (and others)	0.0	1	127	300	1.0

Table 8: Ngbaka inseparability values, first iteration

Iteration 2 allows the learner to unify the prenasalized labiovelar [ŋm gb], whose inseparability rises to 466.47. We learn on this iteration that the labiovelar /ŋm/ occurs overwhelmingly as the first half of [ŋm gb]: 380/385 [ŋm]s appear as part of this longer sequence. It is thus likely that the existence of [ŋmgb] has facilitated unification of [ŋm]. The other sequence unified on this iteration is [n w], with an inseparability of 2.78. On the third iteration, only the residue clusters remain. The learner calculates high inseparability values for these sequences, but it does not unify them due to their low overall frequency. Within the consonant distributions of Ngbaka, there is a difference between [n w], which occurs just 14 times, and the residue clusters, which occur between 1 and 4 times each. The learner detects this difference and reacts appropriately, keeping the low-frequency clusters as clusters. The results for the second and third iterations are summarized in Table 9. As before, clusters that are unified are above the line; the remaining clusters are below it.

sequence	insep (it. 2)	insep (it. 3)	N(C1C2)	N(C1)	N(C2)	p(C1C2)
ŋm gb	466.47	—	380	385	924	0.0
n w	2.78	—	14	325	249	0.0
r h	0.32	85.81	2	127	112	>.1
r t	0.1	27.03	3	127	800	>.1
r k	0.1	26.17	4	127	1469	>.1
r ŋ	0.05	17.67	1	127	198	>.1
r gb	0.04	12.13	2	127	924	>.1
r w (and others)	≤0.04	10.22	1	127	249	>.1

Table 9: Ngbaka inseparability values, second and third iterations

The learner thus succeeds in addressing the challenges posed by Ngbaka. It finds the segment /ŋmgb/ by first unifiying its two subparts [ŋm] and [gb], and then unifying [ŋm gb] on a second iteration. It is able to differentiate complex segments from clusters due to their different frequency: the number of each individual cluster is not significantly different from 0, so the clusters never qualify for unification.

3.2 Prenasalized consonants in Mbay

Mbay (Nilo-Saharan) is described as having voiced prenasalized stops [mb, nd, nɟ, ŋg] (Keegan 1996, 1997). Its segmental inventory, as described by Keegan, is given in Table 10.[7] All nasal-stop sequences are separable in Mbay (in Riehl's 2008 sense), as all subsegments occur independently. The arguments for a complex segment analysis of nasal-stop sequences are mainly phonotactic. One is that they have the same distribution as simplex obstruents: they can occur in syllable-initial but not syllable-final position, where only the sonorants [m, n, ŋ, l, r, j, w] are permitted.

	labial	alveolar	palatal	velar	glottal
stops	p, b, ɓ	t, d, ɗ	ɟ	k, g	
prenas. stops	mb	nd	nɟ	ŋg	
fricatives		s			h
nasals	m	n	(ɲ)	ŋ	
liquids		l, r			
glides	w		j		

Table 10: The consonant inventory of Mbay (Keegan 1997)

While other clusters exist in Mbay, they are licit only intervocalically; word-initially, they are repaired through epenthesis (compare the licit medial clusters in 9f–g to the repaired cluster in 9h). The examples in 9 also illustrate other aspects of Mbay phonotactics: nasal-stop sequences

[7] Keegan (1997:2) specifically notes that the nasal portion of /nɟ/ is not palatal, so we follow this characterization. The palatal nasal has a restricted distribution and is an allophone of /j/ (it appears only word-initially before a nasal vowel), so we do not represent it as a distinct phoneme in our learning data. Keegan treats [ŋ] as a word-final allophone of /ŋg/. As there is no evidence from alternations to this effect, we represent both /ŋ/ and /ŋg/ in the learning data.

occur word-initially and medially, and there is a contrast between a tautomorphemic prenasalized stop [nd] and a heteromorphemic nasal-stop sequence, where the nasal bears tone, 9c–e.

(9) Mbay phonotactics (Keegan 1997:2–11)

a.	dāŋ	'misery'	e.	kùndə́	'millet drink'
b.	nàr	'money'	f.	sèrbétè	'towel' (French *serviette*)
c.	ndà	'hit'	g.	làmpɔ̃̀ɔ̃̀	'taxes' (French *l'impôt*)
d.	ǹda	'he show'	h.	pèlə́r̀	'flower tree' (French *fleur*)

The learner of Mbay again faces a more complex problem than does the learner of Fijian. In Fijian, all consonant sequences were properly analyzed as complex segments. In Mbay, only the nasal-stop ones are, and they must be distinguished from true clusters in order for the learner to match the phonotactic grammar that a phonologist might proffer (i.e. sonorants can be codas, and any singleton can be an onset).

Our corpus for Mbay was a digitized version of Keegan's (1996) dictionary, with 4046 entries. Our learner arrived at the target analysis of the Mbay inventory in one iteration. Figure 2 visualizes inseparability measures for the top 15 of 119 clusters, in descending order (clusters are on the y-axis for readability). The four prenasalized stops fall well above the threshold of 1 (vertical line). The other consonant sequences are close to zero—even on the second iteration, after the nasal-stop sequences have been unified. The plots represent the differences between clusters that qualify for unification based on the Fisher test with round dots, and those that do not with x (in Fig. 2, this is only [n h] on the 2nd iteration).

The calculations for Iteration 1 are presented in more detail in Table 11, which demonstrates the large gap between the inseparability measures of nasal-stop sequences and other clusters.

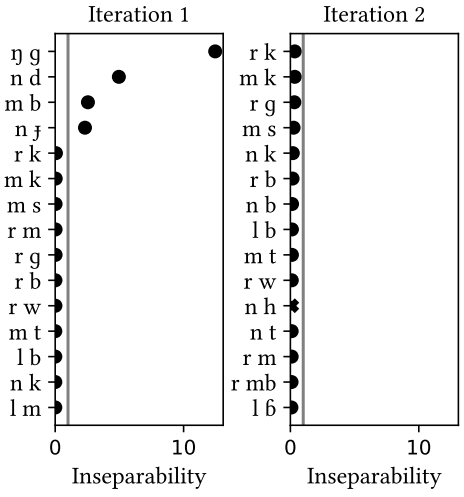

Figure 2: Top 15 inseparability values for various CC sequences at iterations 1 and 2 (Mbay)

	Inseparability	N(C1C2)	N(C1)	N(C2)	p(C1C2)
ŋ g	12.46	579	173	1301	< .001
n d	4.96	352	987	947	< .001
m b	2.55	246	1012	878	< .001
n ɟ	2.32	176	987	505	< .001
r k	0.06	52	1230	1478	< .001
m k	0.04	41	1012	1478	< .001
m s	0.03	22	1012	578	< .001
r m	0.03	31	1230	1912	< .001
r g	0.02	35	1230	1301	< .001
r b	0.02	24	1230	1012	< .001

Table 11: Mbay inseparability at Iteration 1

Mbay differs from Ngbaka in one fundamental way. In Ngbaka, the learner does not unify its remaining clusters because they are too infrequent. In Mbay, by contrast, most clusters are not unified because they are too separable. Most Mbay clusters are frequent enough to qualify for unification, but they are not unified because the segments that compose them combine relatively freely. The difference between these two otherwise similar cases highlights why it is necessary for a cluster to pass checks for both frequency and inseparability before being unified.

Even though Mbay represents a case where complex segments occur along with clusters, the statistical distribution of complex segments still differs from that of consonant clusters: while nasals and stops combine with each other frequently, the combinatoric possibilities are otherwise relatively free. While it is true that Mbay complex segments are more frequent than other consonant sequences (see 8), we will see that this is not a necessary feature: some languages have true clusters that are about as frequent as complex segments. What matters is inseparability.

3.3 Turkish and Hebrew affricates

Here, we sketch two additional cases with clear arguments for complex segments: Turkish and Hebrew affricates. The learner identifies the affricates traditionally posited for the languages, drawing a clear distinction between them and other CC sequences—which in these languages number in the hundreds. Also, unlike the previous cases, affricates in Turkish and especially Hebrew often combine into clusters with other consonants. Iteration here does not result in unwarranted unification.

Turkish has two affricates, [tʃ, dʒ] (Göksel & Kerslake 2004, Kornfilt 2013). Phonotactically, Turkish is a CVC(C) language: CC clusters are allowed word-finally and medially, but not initially. As shown in 10, these generalizations hold only if [tʃ] and [dʒ] are treated as complex segments. In normal colloquial Turkish, loanwords with initial clusters have epenthesis, but [tʃ] and [dʒ] are unaffected (e.g. 'jazz' is [dʒas], not [diʒas], 10c). While [dʒ] is more restricted than [tʃ] (historically, [dʒ] is borrowed), it still patterns more like a segment than a cluster.

(10) Turkish phonotactics (from Göksel & Kerslake 2004:ch.1)

a.	kʰara	'black'	d.	gentʃ	'young'	g.	sitres	'stress' (loan)
b.	tʃene	'chin'	e.	ʃans	'luck' (Fr. *chance*)	h.	kʰiral	'king' (loan, *kral*)
c.	dʒas	'jazz' (loan)	f.	yst	'top'	i.	alarm	'alarm' (loan)

Our corpus was the Turkish Electronic Living Lexicon (65,828 words arranged into paradigms, Inkelas et al. 2000; we got equivalent results using all wordforms vs. citation forms). The learner ran one iteration, unifying [dʒ] (insep. 8.74) and [tʃ] (2.62). The next most inseparable cluster, [n d] (0.36), is nowhere near the threshold. After the affricates were unifed, a total of 362 distinct clusters remained.

Modern Hebrew has one affricate, [ts]. The arguments for this analysis are laid out in Bolozky (1980). Hebrew allows initial clusters but limits them to two consonants (with rare exceptions in loanwords, like [skleʁozis], 11q). With respect to this restriction, [ts] functions as a single segment: witness [btsalim], [tsdaka], 11f–g. Hebrew has also borrowed some words with [t ʃ] and [d ʒ] from English, but their behavior does not clearly motivate a complex segment analysis (Bolozky 1980, Asherov & Bat-El 2019, Asherov & Cohen 2019).

(11) Hebrew phonotactics (Asherov & Bat-El 2019)

a.	kvisa	'laundry'	g.	tsdaka	'charity'	m.	stsena	'scene (loan)'
b.	tkufa	'period'	h.	tʃuva	'answer'	n.	lantʃ	'lunch (loan)'
c.	tsfaʁdea	'frog'	i.	tzuza	'movement'	o.	tʃips	'chips (loan)'
d.	dgima	'sample'	j.	tsvita	'pinch'	p.	dʒins	'jeans (loan)'
e.	psolet	'waste'	k.	tsnim	'toast'	q.	skleʁozis	'sclerosis (loan)'
f.	btsalim	'onions'	l.	tnuva	'yield (n)'	r.	*tʃn, dʒv, etc.	

We tested our learner on the Living Lexicon of Hebrew Nouns (11,599 words, Bolozky & Becker 2006). On iteration 1, the learner identified [ts] (insep. 1.74); the runner-up, [dʒ], was nowhere near the threshold (insep. 0.26). Iteration 2 found no further complex segments, leaving a total of 297 clusters.

3.4 Quechua: the nature of the learning data

We conclude this section with Bolivian Quechua (Parker & Weber 1996, MacEachern 1997, Gallagher 2011, 2013, 2016). Quechua presents clear analytic arguments for affricates, which we introduced in §1: its inventory structure, local phonotactics, and nonlocal restrictions on laryngeal co-occurrence[8] all justify complex segment representations for [tʃ, tʃ', tʃʰ]. Our goal in this section

[8]Since the laryngeal phonotactic restrictions in Quechua are well-studied, we have indirect behavioral evidence that affricates and stops pattern together (Gallagher 2019). Gallagher investigated the co-occurrence restrictions using a nonce word repetition experiment, which demonstrated that Quechua speakers repeat phonotactically legal wugs (e.g. [t'aʎwa, k'apu]) more accurately than wugs with ejectives preceded by [k] or [ʁ], an allophone of /q/. The experiment was not designed to study affricates directly, but the materials include enough stops and affricates for us to test whether they pattern together—is *[katʃ'a] as bad as *[tant'a]? We analyzed Gallagher's data in a model that tests whether affricates are different from stops. We added to Gallagher's model a predictor for whether or not an affricate is present. If there were a difference between affricates and singleton stops, then adding this predictor should improve model fit, and it did not do so for either experiment.The ANOVA model comparison assessed a model

is to use Quechua to answer the question from §2.7: what data supply the right distributions for discovering complex segments? The evidence from Quechua suggests that these representations are learned from type frequencies in morphemes, not phonological words.

For our Bolivian Quechua simulations, we used three kinds of data:

(i) roots (2,479, compiled by Gallagher from Laime Ajacopa 2007),

(ii) morphemes (1,484, from a newspaper corpus, Gouskova & Gallagher 2020),

(iii) phonological words, (10,847, newspaper corpus, Gouskova & Gallagher 2020),

When trained on (i) and (ii), the learner found the target inventory [tʃ, tʃʼ, tʃʰ] in one iteration (see Fig. 3). The inseparability value plot reflects the relative frequencies of [tʃ] (292 occurrences in the roots corpus) vs. the ejective [tʃʼ] (180) and aspirated [tʃʰ] (74). The plain affricate is far more frequent.

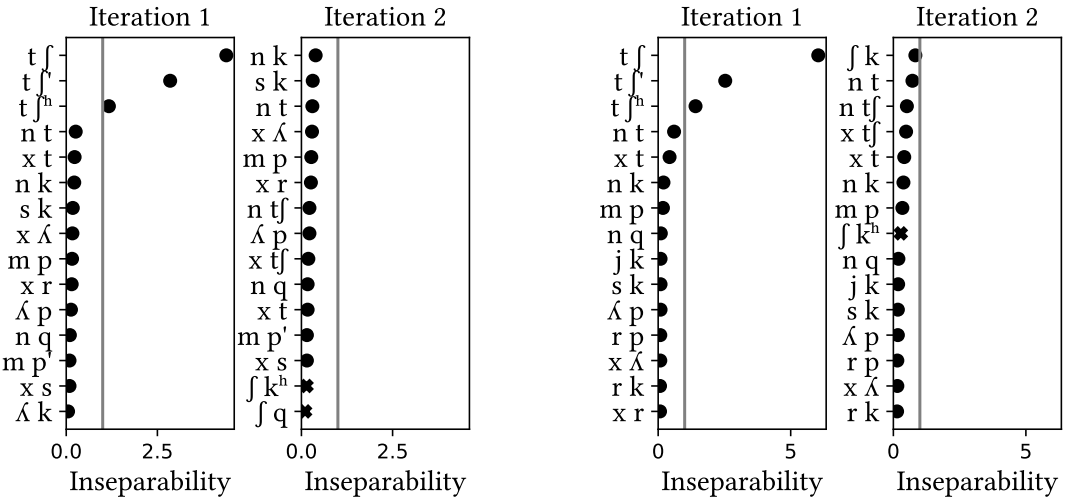

Figure 3: Quechua simulations, roots (left) and morphemes (right)

By contrast, the learner does poorly when trained on Quechua words, (iii). The learner runs nine iterations, unifying [tʃ] and [sq], then [ɲtʃ], [sk], [jk], and [tʃʼ], [rq], and so on. It eventually finds all affricates, but it also unifies all sorts of other sequences. The reasons for this failure become clear when we consider where these "inseparable" clusters occur. First, Quechua has mostly templatic roots, CV(C)CV, but its suffixes are atemplatic and often begin with consonant clusters (e.g. *-sqa* 'nominalizer', *-jku* '1PLURAL.EXCLUSIVE', *-rqa* 'PAST'). Quechua is exclusively suffixing, so when its roots combine with such suffixes, the result is CVC syllables (e.g. [ʎaŋkʼa-rqa-ŋki] 'work-PAST-2SG', [puri-spa] 'walk-GERUND', [hamu-sqa-jki-ta] 'come-PARTITIVE-2SG-ACCUSATIVE'). Second, stop distribution is restricted: (i) ejectives and aspirates (including [tʃʰ, tʃʼ]) do not occur in suffixes, (ii) no stops in codas, (iii) aspirates and ejectives

for *accuracy* as a function of *type* (*control*∼*[k]*∼*[q/ʁ]*) with random by-subject slope for *type* and *affricate*. Adding *affricate* as a fixed effect was not justified by model comparison ($\chi^2(1)=0.23$, p=0.6312). The Akaike Information Criterion (AIC) allows for comparing models with dissimilar structure; on this, too, Gallagher's model with the fixed effect *type* and a by-participant random intercept and slope for *type* was a better fit to the data than the model with *affricate* added to the fixed and random effects (*type* model, AIC=542; *type+affricate* model, AIC=547; lower is better).

cannot follow stops. As a result, [tʃ] is common (3,494 occurrences) and inseparable (5.32), but [tʃʰ, tʃʾ] are less common/inseparable than clusters in common suffixes. Training the learner on morphologically complex words inevitably leads it to unify the wrong things.

We ran a fourth simulation on child-directed speech in the Peruvian dialect (Gelman et al. 2015; one utterance per line, spaces/punctuation removed). The learner found [tʃ, tʃʾ] on iterations 1 and 2, but no other complex segments—quantitatively, close to the right result. But qualitatively, the same pathology arose as in the word simulation: on iteration 3, [tʃʰ] trailed seven other clusters such as [sq, jr], common in suffixes.

This suggests that attending to type frequencies in words or token frequencies in connected speech are the wrong strategies for a language like Quechua. More abstract data are needed. This was not true for other languages, including the agglutinative Turkish—presumably because the affricates in those languages are more evenly distributed among the morphemes. But we do not know what this means for learning: does the Quechua learner follow a different path than the Turkish learner? We return to this issue when we consider Russian in §4.2.

4 Case studies II: Adjudicating between complex segments and clusters

The second set of case studies includes languages where the arguments for complex segments are not clear, regardless of how linguists characterize the inventories. We discuss four such cases here: Latin [kw, gw] (§4.1), Russian [ts, tɕ] (§4.2), English [tʃ, dʒ] (§4.3), and Sundanese (§4.4). Another case in this category is Modern Greek [ts, dz], discussed in §6.4. Unsurprisingly, given the unclear phonological status of these sequences, the learner finds complex segments in some cases and clusters in others.

4.1 Latin [kw, gw]

The consonant inventory of Classical Latin in Table 12 is adapted from McCullagh 2011:84.[9] Our interest is in the sequences [kw] and [gw], whose status as complex segments is marked as questionable.

	labial	dental	alveolar	palatal	velar	labiovelar	glottal
stop	p, b	t, d			k, g	kw, gw (?)	
nasal	m	n			ŋ		
fricative	f		s				h
trill			r				
approximant			l	j		w	

Table 12: Classical Latin consonant inventory

[9]We do not include /pʰ tʰ kʰ z/, as according to McCullagh, these were only attested in Greek loans. We also removed a question mark associated with /ŋ/, as a minimal triplet provided by McCullagh (p. 87: [amniː] 'river' vs. [anːiː] 'year' vs. [aŋniː] 'lamb') suggests it is contrastive.

The arguments for a complex segment analysis of [kw] and [gw] are not convincing, as discussed in detail by Devine & Stephens (1977:ch. 9). One argument is that [kw] and [gw] are the only stop-[w] clusters in Latin (no [pw, tw, dw], etc.). As Devine & Stephens point out, this does not rule out a cluster analysis: 'such rules frequently have odd exceptions which complicate [to] no end the clever flow charts of the phonotacticians, and if *w* is to appear after one stop only, then it is likely that this will be a velar' (1977:90). They cite languages such as Thai, whose only stop-[w] sequences are also dorsal, [k w] and [k^h w], but are analyzed as clusters (on the general preference for labialized dorsals see §5.4). A second argument is that the Roman grammarians treated [kw] and [gw] as segments, so we should too. As Devine & Stephens (1977:ch.4) carefully lay out, however, the segmental status of [kw] and [gw] has likely been debated since late Republican times. A third argument for monosegmental [kw] is that it consistently did not make position in Latin poetry (Devine & Stephens 1977:51–68,McCullagh 2011). This is in contrast to stop-liquid clusters (e.g. *t r*, which sometimes do) and other clusters (e.g. *k t*, which always do). We do not think this proves that [kw] is a segment, since there are many other reasons why it might metrify differently from other clusters.[10] In sum, every argument for treating [kw, gw] as segments is vulnerable to an entirely reasonable counterargument.

We tried several datasets for Classical Latin: a list of 1739 noun paradigms flattened into a list of 12,149 unique forms,[11] Whitaker's online dictionary[12] (84,000 words), and Lewis et al. 1969 (49,725 words). The results were qualitatively the same: our learner found no complex segments. Figure 4 shows inseparability values for the top 15 clusters in each simulation. Neither [k w] nor [g w] are near the threshold; [g w] never makes it into the top 15. The inseparability of [g w] is between 0.01 and 0.07 depending on the simulation. Our results suggest that [k w, g w] were clusters in Classical Latin.

[10] One alternative explanation is that the relevant unit of weight in meter is the interval (Steriade 2012). If so, the different behavior of [k w] tells us is that it was shorter than other clusters. This account could also help explain why stop-liquid clusters made position less frequently than other types of clusters; they may have been shorter (see McCrary 2004 for durational data from Italian, and Steriade 2012 for its potential relevance to meter). Another possibility is that [k w] and [t r] were simply syllabified differently; languages are known to syllabify sequences differently depending on sonority (Vennemann 1988, Gouskova 2004).

[11] https://linguistics.ucla.edu/people/hayes/learning/latin.zip

[12] http://mk270.github.io/whitakers-words/

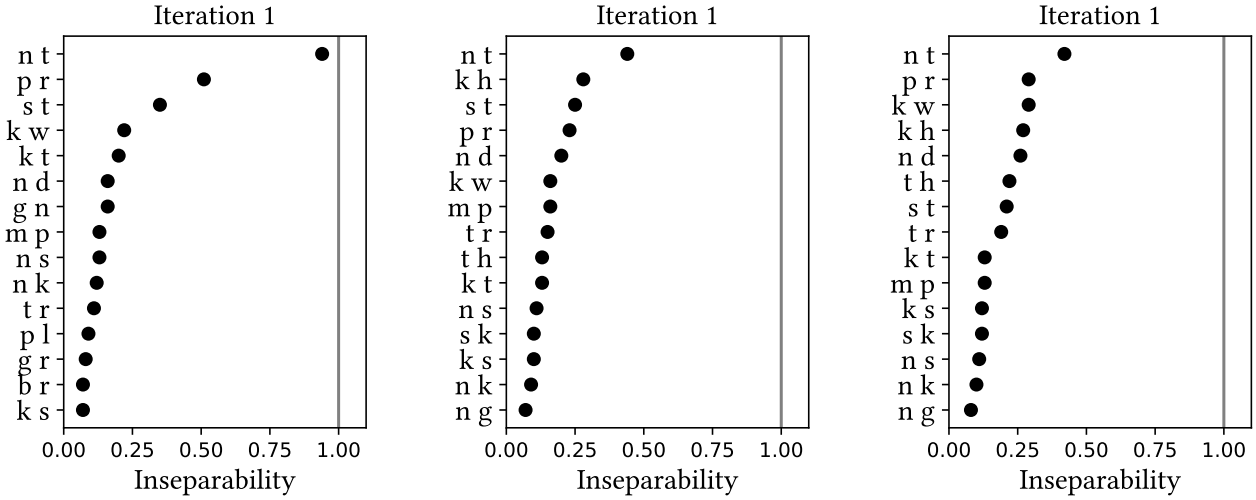

Figure 4: Latin simulations: nouns (left), Whitaker (center), Lewis (right)

In the case studies up to this point, the learner treated many reasonably frequent consonant sequences as complex segments, so it is worth asking whether the learner would insist on finding complex segments even in a language where their motivation is unclear. Latin supplies a sanity check: the learner does not find complex segments in every dataset (see also Modern Greek in §6.4, and French on the project site; French uncontroversially lacks complex segments).

As anticipated in §2.6, Latin is interesting for another reason. The history of the Romance family is well studied, so Latin forms a good baseline for testing the hypothesis about complex segment distributions resulting from sound change. When Latin became Italian, [k g t d] became [tʃ dʒ ts dz] before [i] (Krämer 2007:ch. 3.2.1). We simulated these sound changes by replacing the affected sequences in our Latin data. The learner unified all four sequences in two iterations (in the Lewis simulation, [t s] did not quite pass the threshold).Thus, while Classical Latin might not have had complex segments, its simplex stops were sufficiently frequent in the right environments to become true affricates in daughter languages.

4.2 Russian affricates

As traditionally analyzed, Russian has two affricates: [ts, tɕ]. The phonological arguments for them are lacking, so we ask whether the statistical distributions offer a clearer clue to the learner, and they appear to: our learner identifies [ts, tɕ] in two dictionary datasets. After establishing this result, we revisit the issue of learning data: when the learner is trained on connected speech or morphemes, it unifies [tɕ] but not [ts].

The inventory usually assumed for Russian is in Table 13.[13]

[13]We did not test the learner on palatalized consonants. Russian has contrasts such as [l͡ʲot] 'ice' ∼[l͡jot] 'pours', [ab͡jom] 'volume' ∼ [grʲib͡jom] 'we row', and it is not clear how to transcribe the distinction between Cʲ and Cj. One possibility would be to transcribe the [j]s with different lengths, so [b j] vs. [b jː]; Kochetov 2006:575 finds a difference in articulatory timing. This would likely result in unification of all palatalized consonants as short [j] would be unattested elsewhere.

	labial	dental	(alv-)palatal	retroflex	velar
stops	p, b, p^j, b^j	t, d, t^j, d^j			k, g, k^j, g^j
affricates		ts	tɕ		
fricatives	f, f^j, v, v^j	s, z, s^j, z^j	ɕ:	ʂ, ʐ	x, x^j
nasals	m, m^j	n, n^j			
liquids		l, l^j, r, r^j			
glides			j		

Table 13: Russian contrastive consonants (Padgett 2003, Padgett & Żygis 2007)

The analysis of [ts, tɕ] as affricates is neither questioned nor supported by argumentation in most sources. This could be for two reasons. First, Russian affricates are known reflexes of historical stops (§2.6). Knowledge of this single-segment origin might have biased researchers against questioning the affricates' contemporary status. Second, the affricates might have escaped analytic scrutiny because in Russian orthography, [ts, tɕ] are written with single letters, <ц, ч>.

Trubetzkoy (1939) does supply some arguments.[14] One is phonetic: he suggests that [ts, tɕ] are durationally more similar to simplex segments than to clusters (1939:58). But these intuitions have not (to our knowledge) been supported by experimental research. Second, Trubetzkoy suggests that [ts, tɕ] have the distribution of single segments, since they can occur word-initially. But so can many other consonant-fricative sequences (see 12). Finally, Trubetzkoy's symmetrical inventory heuristic would actually argue against [ts, tɕ]: as Table 13 shows, the inclusion of affricates makes voicing and strident contrasts more gappy. Russian lacks contrastive voiced affricates—[d z, d ʐ] occur only in loanwords, 12n–o. Russian also lacks palatalized and retroflex affricates; [t s^j, t ʂ] (with dental [t]) occur, often at morpheme boundaries, but are uncontroversial clusters, 12r–s.

(12) Russian phonotactics

a.	tsɨna	'price'	k.	kfrantsii	'to France'
b.	tɕuʂ	'nonsense'	l.	pʂɨno	'millet'
c.	v^jetɕir	'evening'	m.	mɕ:en^{j}ijə	'revenge'
d.	r^jetɕ	'speech'	n.	gʐel^j	'Gzhel (place)'
e.	agur^{j}ets	'cucumber'	o.	dʐɨnsɨ	'jeans (Eng.)'
f.	tsv^{j}et	'color'	p.	im^{j}itʂ	'image (Eng.)'
g.	tɕl^{j}en	'member'	q.	dzerʐinsk^{j}ij	'Dzerzhinsky (Polish)'
h.	ksv^{j}in^{j}je	'to a pig'	r.	ot-ʂit^j	'to ditch'
i.	kxar^{j}ku	'to a hamster'	s.	ot-s^{j}ejat^j	'to weed out'
j.	kxval^{j}e	'to praise'			

We can supply (and refute) one more argument: affricates alternate with segments, as in

[14] Another exception is Halle 1959. His argument for [ts] is subtle, and is based on vowel-zero alternations, [met_l-a]∼[met^{j}ol] 'broom (nom.sg/gen.pl). No morphemes exhibiting these alternations end in CC; if [ts] were a cluster, alternations in [ofts-a]∼[ov^{j}ets] 'sheep (nom.sg∼gen.pl)' would be the only exceptions. The alternations are lexically restricted, and the same suffix, [-ets] is the main source of [ts]-final examples (see Gouskova 2012, Becker & Gouskova 2016). While we are sympathetic towards this argument, it seems like the stakes are low; the learner who misses the segmental status of [ts] could just conclude that [-ets] is an exception within exceptions.

[krʲuk] ‘hook’ ∼ [krʲutɕ-ja] ‘hooks’, [durak] ‘fool’ ∼[durats-kij] ‘foolish’. The problem with this argument is that Russian segments also alternate with uncontroversial clusters (e.g. [pabed-il] ‘he won PERFECTIVE’ ∼ [pabezd-al] ‘he won IMPERFECTIVE’). If the learner uses alternations as a cue for unifying some clusters into segments, then it still needs some heuristics to decide which clusters to unify.

In short, it is not obvious to us that an analyst without preconceptions about Russian would posit the particular affricates of the traditional analyses.

We tested two digital dictionaries: Zaliznjak 1977 (93,392 words) and Tikhonov 1996 (101,531 words, reported here). The learner unified [tɕ] in the first iteration and [ts] in the second. The results are shown graphically in Figure 5:

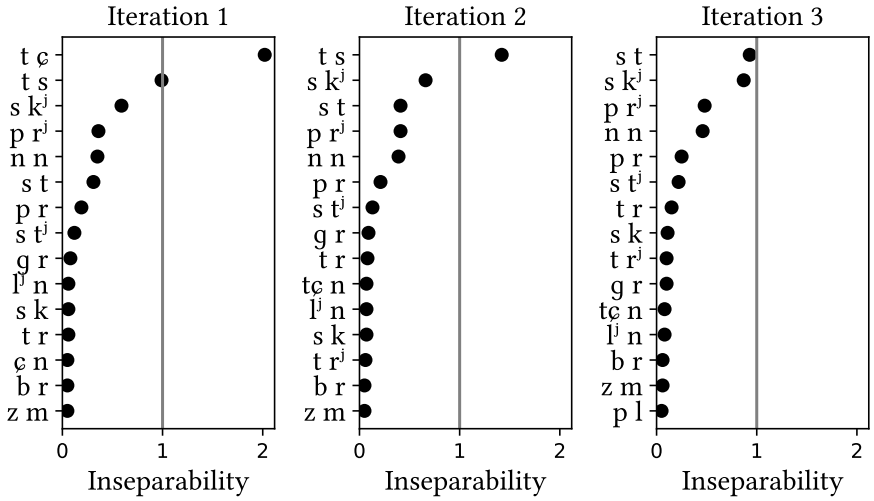

Figure 5: Inseparability for Russian CC sequences

This result suggests that the best argument for the single-consonant analysis of Russian [ts, tɕ] is their distribution. While the result is replicated on two training data sets, there are reasons to be skeptical of it. First, the frequency of [ts] might be artificially inflated in dictionaries by infinitival forms of verbs, which often end in [tsa], with t-s straddling a morpheme boundary. We therefore tried two alternative data sets for Russian: first, we created a connected speech corpus, and second, we tested a tokenized morphologically segmented corpus. The connected speech corpus was created by pasting together 12 Russian novels. The words were automatically transcribed, and connected speech rules were applied; spaces and punctuation marks were then removed. The resulting corpus had almost 6 million characters in it. The second corpus was created on the basis of Tikhonov 2002, with some manual correction. It was transcribed and tokenized, yielding 18,707 affix and root allomorphs (we did not attempt to unify them into unique URs). Trained on either connected speech or morphemes, the learner identified [tɕ] but no other complex segments. In connected speech, inseparability for [tɕ] was 1.89; in morphemes, 2.88. There were qualitative differences between simulations: In morphemes, [t s] (0.47) followed [s t] as the second runner-up (0.51); in connected speech, [t s] was 7th in the last iteration (insep=0.2), trailing clusters in common morphemes such as [p rʲ].

The divergence between the dictionary simulations and morpheme and connected speech ones is difficult to interpret. There is no clear evidence that [ts] is an affricate in Russian. This is equally true for [tɕ], but the latter is much easier to unify in a variety of data. We could certainly question the quality of the mock connected speech dataset and the morphemic lexicon, and trust the dictionary simulations, which find both affricates. While this runs counter to the results of the Quechua simulations, there are significant morpho-phonological differences between Quechua and Russian. Russian allomorphy and fusional morphology make it difficult to segment, presumably both for linguists and for learners. By comparison, a Quechua learner might have an easier time arriving at a mental lexicon of morphemes early on than a Russian learner. Until better evidence comes to light, we simply do not know what the status of these sequences is.

4.3 English affricates

We wanted to test English because its phonology has been studied in more detail than any other language, and the phonotactics are well-understood (Jones 1918, Scholes 1966, Chomsky & Halle 1968, Kahn 1976, Selkirk 1982, Borowsky 1986, Moreton 2002, Daland et al. 2011). Just as in Russian, the phonotactic arguments for the traditional analysis of the inventory are problematic, but our learner does identify the two affricates [tʃ, dʒ] when given nuanced evidence.

The traditional analysis of English is that [tʃ, dʒ] are complex segments but [t s, d z] are clusters (Jakobson et al. 1952:43, Chomsky & Halle 1968:223; cf. Jones 1918). This rests on a phonotactic argument: [t s, d z] do not occur word-initially (aside from careful pronunciations of loanwords such as *tsunami*; e.g. Ladefoged 1996). Under this analysis, [t s] and [d z] cannot occur word-initially because stop-fricative clusters are not allowed in initial position. But this is unsatisfying, as English phonotactic constraints include bans on singletons in word-initial position, such as [ŋ]. The same ban could apply to [ts, dz], were they single segments. The analysis also has difficulty explaining why [tʃ] cannot combine with other consonants word-initially. English [tʃ] patterns differently than both [ʃ] and [t], which can combine with approximants: [ʃ w, ʃ l, t w] but not *[tʃ w, *tʃ l] (cf. Hebrew [ts], which clusters like simplex segments). It is not clear that an analyst without preconceptions would arrive at the traditional analysis of English on the basis of phonotactics alone—and it is even less clear what evidence the English learner would use.[15]

We tried two corpora: Celex (Baayen et al. 1993, 72,969 words) and the Carnegie Mellon Dictionary (version of Hayes & White 2013). We describe the Celex runs here, though we got the same qualitative results on CMU. We tested two versions of the corpus. First, we transcribed the postalveolar affricates narrowly, with retracted "allophones", [c ʃ] and [ɟ ʒ] (the retracted diacritics [t̠, d̠] are more appropriate but harder to see). Celex indicates morpheme boundaries (as syllabification) in its transcriptions, so we could even differentiate acoustically distinct sequences: [t ʃ] is alveolar-postalveolar in *courtship,* but postalveolar-postalveolar [c ʃ] in *ketchup.* When trained on these transcriptions, our learner identifies [c ʃ, ɟ ʒ] as affricates on the first iteration and finds no other complex segments on the second iteration. In both iterations, [t s] is well below the threshold (Fig. 6).

[15] A reviewer suggests that experiments such as Pig Latin (Barlow 2001) could be used to probe the status of English [ts, dz, tʃ, dʒ]: [tʃ] should never be split in *church,* but if "onset splitters" ([spun]→[pun-seɪ]) produce *tsunami* as [unɑmi-tseɪ], that would show [ts] to be an affricate. Barlow (2001) reports that speakers do split *church*, but orthographically: *shoe*→[hu-seɪ], *church*→[hɚtʃ-tʃeɪ], *thumb*→[hʌm-θeɪ]. This casts doubt on Pig Latin as a test of phonological knowledge— it is too metalinguistic and tied to literacy (cf. Cowan 1989, Hester & Hodson 2004).

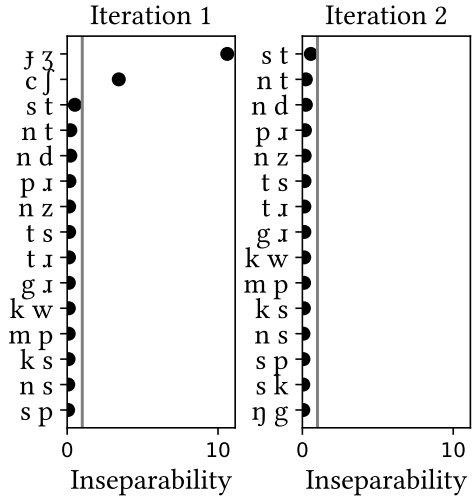

Figure 6: English inseparability measures, affricates transcribed narrowly

This result is unsurprising, as the setup is rigged in favor of finding the affricates; [c] and [ɟ] only occur as part of [c ʃ] and [ɟ ʒ]. The difference in their inseparability measures is due to the differing frequencies of singleton [ʃ] and [ʒ]; [ʒ] is far rarer (see Table 14). Note also that the cross-morpheme and nonhomorganic [t ʃ] has an inseparability of 0 ([d ʒ] is not included as no such sequences exist). No other clusters approach the inseparability threshold on either iteration, which indicates that aside from the affricates, consonants in English combine relatively freely.

	inseparablity	N(C1C2)	N(C1)	N(C2)	p(C1C2)
ɟ ʒ	10.61	4002	4002	4332	< .001
c ʃ	3.42	2730	2730	9162	< .001
t ʃ	0.00	35	36312	9162	< .001

Table 14: English inseparability calculations for Iteration 1 under narrow transcriptions

When the learner is trained on broadly transcribed data, [d ʒ] but not [t ʃ] qualifies for unification. This difference between the two sequences is again due to the overall rarity of [ʒ] (see Table 15). Both [t] and [ʃ] are fairly frequent, so the inseparability of [t ʃ] is below 1. As was the case for the narrowly transcribed simulations, no further clusters qualify for unification on the second iteration.

	inseparablity	N(C1C2)	N(C1)	N(C2)	p(C1C2)
d ʒ	1.64	4002	25859	4332	< .001
t ʃ	0.25	2765	39042	9162	< .001

Table 15: English inseparability calculations for Iteration 1 under broad transcriptions

To conclude, the quantitative support for the affricate analysis of English [tʃ] is not strong. This is because [t ʃ] is not frequent enough to counterbalance the individual frequencies of [t]

and [ʃ], so the learner fails to unify it without being given more detailed phonetic information. Regardless, our learner is clear on the status of [t s] in English: it is a cluster, not an affricate.

4.4 New predictions: Sundanese nasal-stop sequences

We next describe a case where our learner's posited segment inventory diverges from the inventory proposed by analysts. Only one of the languages we have investigated—Sundanese—clearly falls into this group. Sundanese nasal-stop sequences are occasionally characterized as complex segments (Blust 1997:170), but it is not clear that there is any evidence for treating them as such. While none of the descriptive work on the language (Robins 1957, 1959, Cohn 1992) explicitly discusses the question of segmenthood, there are hints throughout that these authors assume they are clusters. Robins provides a CV representation of [sunda] as CVCCV and [ŋimpi] as CVCCV (1957:89), and refers to them as sequences (his fn. 1). Cohn does not include them in her posited inventory, and describes nasal-stop sequences as split across a syllable boundary (1992:205). Nonetheless we were interested in testing our learner on Sundanese, as different claims have been made regarding the segmental status of its nasal-stop sequences.

The uncontroversial consonants of Sundanese are in Table 16. Following Cohn (1992:205), we treat /s/ as palatal and /w/ as labial. The distribution of [ʔ] is largely predictable (see Robins 1959:341–342) so, again following Cohn, it is in parentheses.

	labial	coronal	palatal	velar	glottal
stop	p, b	t, d	c, ɟ	k, g	(ʔ)
nasal	m	n	ɲ	ŋ	
fricative			s		
liquid		l, r			
glide	w		j		h

Table 16: Sundanese consonant inventory following Cohn 1992

Cohn (1992:205) describes the phonotactics of Sundanese roots as follows. Any consonant can occur as a singleton onset. A word-final coda can be any consonant except [c] and [ɟ]. More relevant here are the constraints on clusters: complex onsets are infrequent (but stop-liquid onsets do occur word-medially), and while coda-onset combinations usually consist of homorganic nasal-stop sequences, the medial coda slot can be occupied by /r/ or another consonant as well.

We trained our learner on Lembaga Basa & Sastra Sunda 1985, a monolingual Sundanese dictionary (16,327 headwords entered manually). In addition to the segments in Table 16, the dictionary included words that contained [f], [v], and [z] (likely loans, like *afghanistan*); these segments were added to the feature table and assigned the appropriate distinctive features. The only way in which our transcriptions deviated from the dictionary's is that all palatal nasal-stop sequences were transcribed with /ɲ/ (rather than the dictionary's *n*), in accordance with Cohn's observation that medial nasal-stop sequences are homorganic.

Our learner found 223 distinct CC sequences on the first iteration, and unified the following seven sequences: [ɲc, nd, ɲɟ, mb, mp, nt, ŋk]. On the second iteration, the learner unifies the only remaining nasal-stop sequence, [ŋg]. On the third iteration, no sequence passes the threshold of

1: [ŋ s] rose to 0.44, and all the other sequences are lower. The overall results are summarized in Figure 7.

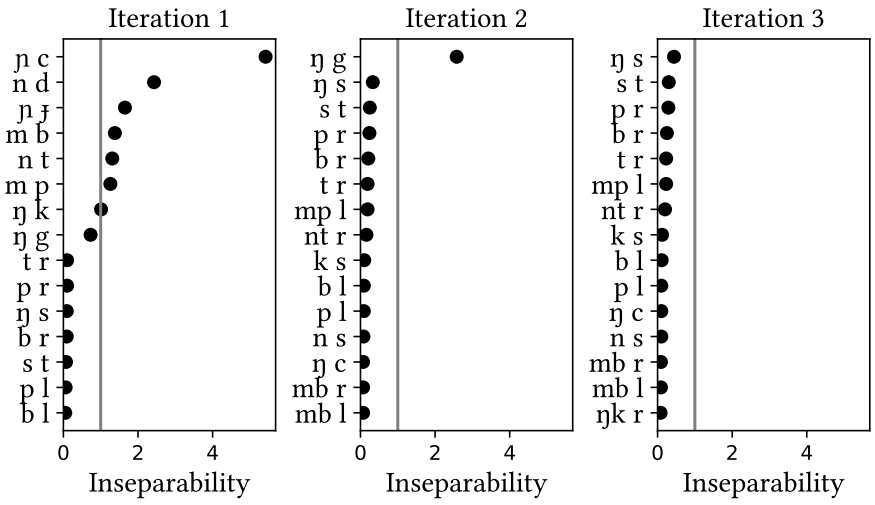

Figure 7: Sundanese simulation

Our learner finds matched sets of voiced and voiceless prenasalized stops at all places of articulation. This result would require characterizing Sundanese as having a phonotactic ban on prenasalized stops in initial position (Cohn and Riehl Cohn & Riehl 2016:5), but such phonotactics can hold of individual segments (e.g. English [ŋ]). The learner's conclusion thus mirrors descriptions that treat the voiced series as complex segments, but goes beyond these descriptions by analyzing the voiceless nasal-stop sequences as segments as well.

This latter point is worth addressing further, in light of Riehl's (2008:52–55) claim that prenasalized voiceless stops (NTs) do not exist. One argument is that NTs are rare. The other is that languages allowing NT sequences necessarily have voiceless stops in their inventory, so the sequences are always separable. This latter observation has been contested: Stanton 2017 notes that Makaa (among other languages) is described as having voiceless prenasalized /mp/ but not /p/ (see Heath 2003). This means that /mp/ is inseparable and would necessarily be analyzed as unary under Riehl's criteria. Regarding the first argument, we endorse Riehl's (2008:53–54) speculation that 'the presumed dispreference for [NT] sequences in general [. . .] combined with the relatively small number of languages that contain prenasalized segments of any kind, results in their rarity' (see Hayes & Stivers 1996, Pater 1999, and our §5.4). As we explain in the next section, under our analysis, the typology of complex segments is predicted to mirror the typology of the same-phone clusters.

Our learner's analysis of Sundanese allows us to make sense of Cohn and Riehl's observation that 'the distribution of NDs completely parallels that of NTs' (2016:37). This observation supplies an argument against analyses that accord only NDs segmental status. But if both types of nasal-stop sequences are in fact complex segments, then the observed parallels in their distribution are less surprising.

5 Typology

Coupled with additional assumptions, our proposal makes predictions for the typology of complex segments. We focus on generalizations about their size, as these have been addressed by other proposals (§5.1), and the typology of complex segment size and cluster size is well understood (§5.2). Section 5.3 discusses Shona, which is typologically unusual in allowing four-part complex segments. Section 5.4 briefly discusses several generalizations about the composition of complex segments.

5.1 Theories of limitations on complex segment size

Typologically, complex segments often have two subparts (mb, ts), less commonly three (ɲdʒ), and rarely four (ɲdʒw). Some proposals capture this by stipulating limits on representation. Aperture theory (Steriade 1993) proposes that complex segments have maximally two positions to which features can dock. Under this proposal, it is possible to represent a segment like [mb] or [ts], but not a segment like [tʃkw], which would necessitate at least three docking sites. The idea is then that complex segments consisting of more than two sequentially ordered nodes are representationally impossible, or excluded from the learner's hypothesis space. Q theory (Inkelas & Shih 2013, 2016, Garvin et al. 2018, Shih & Inkelas 2019a,b, cf. Schwarz et al. 2019) imposes similar limitations on the size of complex segments. In Q theory, each segment consists of sequenced subsegments, and most work assumes a maximum of three: 'Q theory makes the strong prediction that a canonical segment can have up to three, but no more than three, featurally distinct and uniform phases' (Shih & Inkelas 2019a:3).

These stipulations offer no independent reason why a complex segment should be limited to two or three subparts. Our theory of complex segments, by contrast, provides a potential explanation. If complex segments are clusters unified due to their statistical distributions, then large complex segments must be rare because large clusters are rare, both within and across languages. This generalization about clusters is well-established in typological research, as we show next.

5.2 Rarity of long consonant clusters

Typologically, the bigger the consonant cluster, the less common it is. Gordon's (2016) study of syllable structure in a sample of 97 languages gives us some idea of the maximum number of consonants that syllables can accommodate, cross-linguistically. We can use his results to estimate the maximum cluster size allowed across these languages, assuming no constraints on combination. In a language that allows CC onsets and CC codas, for example, the maximum cluster size will be four (VCC.CCV). The number of languages per predicted maximum cluster size, given Gordon's survey, is in Table 17. A minority (30/97) are predicted to allow clusters with 4 or more members.

Max cluster size	1	2	3	4	5	6
No. of languages	8	34	25	16	6	8

Table 17: Predicted maximum cluster size, calculated from Gordon 2016:91

Large consonant clusters are rare not only cross-linguistically but also within languages. Even for the 30 languages in Table 17 where the predicted maximum cluster size is 4–6, the learner would probably rarely see clusters of this length. For example, our Russian corpus of 101,531 words contains 289,830 intervocalic consonant sequences (counting affricates as complex segments). Russian has clusters up to five Cs, but they only occur 31 times in our corpus. As Table 18 makes clear, single consonants and CC clusters are far more common.

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%

Table 18: Frequency of intervocalic consonant sequences in Russian

A corpus study of 16 languages by Rousset (2004) makes the same point: there is likely an inverse correlation between cluster length and frequency of attestation. Kannada, for example, allows CC onsets and codas, meaning the maximum cluster length in this language is four. But these four-consonant clusters are likely infrequent: based on the frequencies of syllables with complex onsets and codas, four-consonant clusters are expected to constitute only .001% of all intervocalic consonant clusters.[16] The rest of the languages in Rousset's study make the same point; see her p. 116 for details.

5.3 Shona: where long complex segments are motivated

Our approach predicts that a language could have four-part, five-part or longer segments if clusters of this length qualify for unification. This has been argued to be the case in Shona (Doke 1931, Fortune 1980, Maddieson 1990, Kadenge 2010, Mudzingwa 2010). We focus on the Zezuru dialect as it is among the best-described. Its simplex consonants are in Table 19.

	labial	alveolar	postalveolar	whistled	velar	glottal
stops	p, b, bʱ	t, d, dʱ			k, g	
fricatives	f, v, vʱ	s, z	ʃ, ʒ	ʂ, ʐ		ɦ
nasals	m, mʱ	n, nʱ	ɲ		ŋ	
liquids		r				
glides	w, ʊ		j			

Table 19: Zezuru: simplex consonants (Fortune 1980)

According to Fortune (1980) and others, the basic phones can combine into affricates [pf, bv, ts, dz, tʃ, dʒ, tʂ, dʐ], prenasalized consonants [mb, nd, nz, ŋg, . . .], and velarized consonants [tw, dw,

[16] The frequencies are: 75.52% CV, 0.38% CCV, 3.43% V, 2.45% VC, 0.03% VCC, 17.91% CVC, and 0.27%. The probability of a four-consonant sequence was calculated by adding the probability of VCC.CCV to the probability of CVCC.CCV, as these are the two ways of creating a four-consonant cluster.

sw, ʃw, ɲw, rw, mw…]. Zezuru also has complex coronal-velar and labial-coronal segments: three-part segments like [dʒg, tʃk, mbʒ], and four-part segments like [dʒgw, tʃkw]. Phonotactically, Zezuru is (C)V (Kadenge 2010): complex segments occur both initially and medially. Evidence for this analysis of Zezuru phonotactics comes from loanword adaptation, where consonant clusters such as [g l], [p r] are broken up by epenthesis (Maddieson 1990:27; as in [ma-girazi] 'glasses (English)', [mu-puraŋga] 'gum-tree (Portuguese *prancha*)').

To see if our learner finds larger complex segments, we trained it on an electronic dictionary (15,830 entries, Chimhundu 1996, transcribed from orthography following Fortune 1980). Over five iterations, our learner finds many complex segments (41 total). The counts for four- and five-consonant sequences are in Table 20, shown as they appeared at the point of unification in iterations 3, 4, and 5. The learner unifies all but [ɲ dʒgw], which passes the inseparability threshold (1.61) but fails the Fisher's Exact test.

Sequence	N(C1C2)	p(C1C2)
[ɲdʒ g]	42	0.0
[ts kw]	37	0.0
[dz gw]	26	0.0
[nz gw]	25	00
[tʃ kw]	12	0.0
[dʒ gw]	6	0.029
[ɲ dʒgw]	1	1.0

Table 20: Counts for Zezuru four- and five-part segments

Zezuru Shona illustrates two points. First, our learner has no trouble finding four-part segments when they are motivated by the data; this would be impossible for a learner hampered by the representational assumptions of aperture or Q theory (if limited to three subsegments). Second, the likely reason why five-part and longer complex segments are not attested is because five-consonant sequences are rare, even in languages like Russian and Shona where they are in principle licit.

5.4 Other predictions: composition

Our proposal might also explain other aspects of the typology of complex segments. In particular, there are indications that complex segments and clusters are similar not only in size but in composition. This follows if (as we assume for present purposes) the constraints that hold of the internal content of these sequences are the same, regardless of whether they have been unified or left as clusters.

Two links between the composition of clusters and complex segments has already been mentioned: (1) there is an affinity between dorsals and [w] (§4.1), and (2) there is a dispreference for voiceless nasal-stop sequences (§4.4). We discuss those in more detail here.

The dorsal-[w] affinity is part of a broader pattern of dorsal-labial interactions in clusters and complex segments (Ohala & Lorentz 1977). Languages with labialized consonants often have a gap of precisely the same combinations that are ruled out as clusters in other languages. Tswana, a close relative of Shona, has a series of complex labialized segments including [xw], [ŋw], [kxw],

[sw], etc. Labialization is contrastive on all dorsals, some coronals, but not labials—Tswana has [p] but not [pw] (Tlale 2005). In this, Tswana differs from Shona, which does have [bw], [mw], etc. Tswana is the complex segment analog of English, whose word-initial stop-[w] clusters are dorsal or coronal (*queen, tweak*) but not labial (Selkirk 1982, Moreton 2002). These patterns follow if a single set of constraints governs combinations of various places of articulation with a [w]-like gesture, regardless of whether the sequences are analyzed as complex segments or clusters.

There is likewise a well-documented typological dispreference against nasal-voiceless-stop (NT) clusters (Pater 1999, Hayes & Stivers 1996, et seq.). For segments, this dispreference manifests in the rarity of voiceless prenasalized stops. Maddieson & Ladefoged (1993:256) note that only 8 languages in UPSID have NT stops (compared to 55 with some kind of prenasalized consonant). Our proposal can make sense of these parallels between NT clusters and NT complex segments if the same constraints on postnasal voicing govern both. Moreover, if NT clusters are rarer than ND ones—either cross-linguistically or within a language[17]—we would expect NT to be unified less frequently.

Other links between the typologies of prenasalized stops and nasal-stop clusters might be similarly explained. For example, the vast majority of prenasalized consonants are homorganic, which can be linked to the common requirement that nasals assimilate in place to following consonants (Mohanan 1982, Ito 1986, Padgett 1995b, and many others). This requirement appears to hold as a statistical trend even in languages that allow heterorganic nasal-stop sequences. In Yindjibarndi, Wordick's (1982) lexicon contains 574 homorganic nasal-stop clusters and 160 heterorganic clusters (Stanton 2019). In Wargamay, Dixon (1981:23) reports that homorganic nasal-stop clusters are four times more common than heterorganic ones. Russian also has more homorganic clusters such as [n t, n t^{j}] (2,386 occurrences in Tikhonov 2002) and [m p, m p^{j}] (548) than heterorganic ones such as [m k, m k^{j}] (197).

Constraints on consonant sequencing could explain some other generalizations about the typology of complex segments. For example, [nd] and [kw] are fairly frequent in the inventories of the worlds' languages, but [nl] is—to our knowledge—unattested (Maddieson & Ladefoged 1993:253–254). This would follow if [nl] were a rarer cluster than [nd] and [kw]. Exploring these links rigorously requires quantitative typological research, which has not been undertaken systematically. But we predict that such research should reveal the composition of complex segments and clusters to be similar. Thus, our proposal allows us to begin to answer a broader question (previously addressed by Herbert 1986, Steriade 1993, a.o.): why are only certain combinations of consonants attested as complex segments?

6 Alternatives

We discuss four alternatives. First (§6.1), we defend calculating probabilities over segments rather than natural classes. In (§6.2), we consider a model which divides utterances into segments top-down, instead of unifying bottom-up. In (§6.3), we consider learning complex segments in tandem with phonotactics. Finally (§6.4), we discuss phonetics as a strategy for identifying complex segments.

[17] These statistical trends do not have to hold in any given language, of course. In our English corpus, [n t] and [m p] are both more common and more inseparable than [n d] and [m b] respectively. But English is complicated: [m p] is allowed word-finally, but [m b] is not (Kaplan 2007).

6.1 Inseparability over natural classes

We treat discovering complex segments as a problem for a learner that already has segments. The model unifies consonants (defined by [-syllabic]), not natural class-based bigrams. But a natural-class-based approach has appeal.[18] First, it aligns with other statistical phonology models, which assume that learners generalize over features and natural classes (Albright & Hayes 2003, Frisch et al. 2004, Hayes & Wilson 2008, Albright 2009, Adriaans & Kager 2010, Gouskova & Gallagher 2020). In this view, a segment is a natural class with one member. Second, and more importantly, learning over natural classes would allow the learner to discover generalizations: thus, English affricates [tʃ, dʒ] are [-son,-cont][+strid, -anterior]. A natural class learner could find this bigram in one iteration.

To explore this alternative, we calculated inseparability over natural classes defined by feature charts. As in our model, we restricted the class-based learner's calculations to sequences of natural classes that contained only consonants. We did not iterate the learning procedure, for reasons explained shortly. Substantively, this learner misses generalizations in some languages, and posits wrong ones in languages with gapped complex segment inventories.

No clear threshold. Unlike our learner, the natural class alternative has no clear threshold identifying sequences for unification. This is due to the combinatorics of natural classes (Hayes & Wilson 2008, Gouskova & Gallagher 2020). The purpose of natural classes is to group segments in various ways: [m] can pattern with labial sonorants, noncontinuants, etc. Correspondingly, there are far more natural classes than segments. Table 21 shows comparison calculations for several languages. The "Nat. Cl." column shows the total number of classes (vowels and consonants). "CC 2grams" counts [-syll] *segment bigrams* attested in the data. The next column counts these same bigrams as *natural classes*; for example, [m b] is counted multiple times as [+nas][-son], [+lab][+voice, +lab, -cont], etc. The last column shows the maximal inseparability values calculated over natural classes.

Language	Segs	Nat. Cl.	CC 2grams	Nat. Cl. 2grams	Max N.Cl. insep.
English (Celex, broad)	36	208	340	8,751	0.0040
English (Celex, narrow)	38	223	371	11,497	0.0220
Fijian	25	164	6	4,136	0.0368
Greek	30	133	181	8,052	0.0036
Hebrew	31	232	272	22,579	0.0013
Latin (Whitaker)	32	159	178	12,463	0.0017
Ngbaka	33	199	28	5,720	0.0086
Quechua (roots)	33	143	124	6,182	0.0200
Russian (Zaliznjak)	41	256	597	45,059	0.0017
Turkish	38	198	315	8,630	0.0126

Table 21: Segment vs. natural classes, by language (first iteration)

Calculating over thousands, as opposed to dozens or hundreds of bigrams, reduces the probability of each bigram so much that inseparability values never approach 1. This makes the notion

[18] We would like to thank Donca Steriade for pressing us on this point.

of a threshold no longer tenable. Furthermore, the inseparability values are not comparable between languages. By contrast, our learner uses simple reasoning that works within and across languages.

No obvious negative results. The lack of a clear threshold makes it difficult to know when the learner yields a negative result. Take Latin, Greek, and Russian vs. Hebrew. Our learner found no unifiable sequences in Latin or Greek, but identified affricates in Russian and Hebrew; the phonotactic arguments are strongest in Hebrew.

The natural-class-based calculations make no obvious cut here. The most inseparable sequence in Latin [+cons, -long][-long], which covers nongeminate consonant clusters. Its inseparability value, 0.0017, is the same as Russian's [tɕ]. Compare this to Greek, whose most inseparable sequence is [+cons,-nas,-syll][+cons], insep=0.0036. This is more than double the value in Russian/Latin. In Hebrew, the most inseparable natural class bigram is [ts], insep=0.0013 (lower than Greek). Calculations over natural classes offer no guidance to how Greek, Latin, Russian, and Hebrew should be treated—whether to iterate, how to stop, and what criterion to use.

Missing and overly broad generalizations. The natural class learner misses generalizations even when present. Consider Fijian. On iteration 1, our segment-based learner unifies the prenasalized stops and then the affricates (§2.2). But as a natural class (Table 22), prenasalized stops [+nas][+voice,-cont,-son] have an inseparability value lower than [ŋg], [mb]. Affricates are also missed. This odd outcome arises because the inseparability calculation has as its denominator the frequency of the sequence's parts, so lumping all nasals and all stops into one calculation dilutes the inseparability of all prenasalized stops. Since each segment is also a natural class, the inseparability of a sequence with less frequent members (e.g. [ŋ, g]) will be higher when it is considered on its own.

nat. class bigram	as segments	Insep	N(Cl1 Cl2)
[+dor,+nas] [+dor,+voice,-son]	[ŋ g]	0.0368	1026
[+lab, +nas] [+lab, +voice, -cont, -son]	[m b]	0.0301	2328
[+nas] [+voice, -cont, -son]	[m|n|ŋ b|d|g]	0.0295	5866
[+ant, +nas] [+ant, +cor, +voice, -cont, -son]	[n d]	0.0269	2512

Table 22: Top most inseparable sequences in Fijian, as natural classes

In one pathological case, the natural class learner finds too broad a generalization, even though the target inventory is gapped and cannot be captured via a natural class: Russian. Our segment-based learner identified [tɕ] and then [ts], in two iterations. The natural class-based learner also identifies [tɕ], but it is tied on inseparability with [t][ɕ|s|sʲ|ʂ] (Table 23). This would unify [tsʲ] and [tʂ], which mostly occur at morpheme boundaries (recall 12) and are clusters in any mainstream analysis. The learner cannot help but overshoot here, because a natural class definition of affricates in Russian is hopeless: the inventory is gapped. A gapped inventory requires a segment-level analysis.

nat. class bigram	as segments	Insep	N(Cl1 Cl2)
[+back,-strid,-voice] [-ant,-back,-voice]	[t ɕ]	0.0017	12900
[+back,-strid,-voice] [+strid,-ant,-back]	[t ɕ|ʑ]	0.0017	12900
[+back,-strid,-voice] [+strid,-voice]	[t s|sʲ|ɕ|ʂ]	0.0017	28049
[+back,-strid,-voice] [+cont,-voice]	[t f|fʲ|s|sʲ|x|xʲ|ɕ|ʂ]	0.0015	28171
[+back,-strid] [+strid,-ant,-back]	[d|t ɕ|ʑ]	0.0014	12905

Table 23: Top most inseparable sequences in Russian, as natural classes

We should emphasize that the natural class calculations often look qualitatively similar to segmental ones—mainly because each segment is a natural class, and bigrams such as Fijian [ŋg] are often most inseparable. In Turkish and English, [dʒ] is most inseparable counted as classes and segments. The learner even occasionally finds the right class-based generalization: in Quechua roots, the topmost sequence is [t ʃ|ʃʰ|ʃʼ], precisely the right result. But the success in Quechua is nothing to celebrate, given how often this learner misses class-based generalizations.

6.2 Top-down segmentation

We are sometimes asked why our learner unifies segments bottom-up. The alternative is segmenting top-down, starting with holistic utterances and then dividing them into segments, leaving some sequences as complex. This is desirable because any model of acquisition must solve the segmentation problem; it is controversial whether infants represent the speech stream as separate phones (see Phillips & Pearl 2015). A more realistic model might proceed in the opposite direction from ours.

We did not implement a top-down model, but we considered an alternative that uses the inverse of our math and does not presuppose the vowel-consonant distinction. In this model, the most separable sequences are divided into segments, and the inseparable ones would be left as complex—undersegmented, not unified. The main problem in this model is similar to the natural class alternative: because there are more bigram types to consider, there is no longer a clear threshold. The numbers look similar in languages that have complex segments and in ones that do not. Moreover, target CC sequences can be more separable than some CV sequences (e.g. in Fijian, [nr] would be segmented before [ka, βa, ta], and in Hebrew, [ts] would be segmented before [ut]). At the very least, then, a top-down segmenter has to treat consonant sequences as special–and then the model is close to being a notational variant of ours.

6.3 Phonotactics

An alternative to our inseparability heuristic is that learners use phonotactics to decide which sequences are complex segments (as hypothesized explicitly by Herbert 1986). Suppose the Fijian learner tries learning phonotactics assuming clusters, [m b] and then tries complex segments, [mb]—and evaluates resulting grammars for improved fit. We show that this strategy fails because phonotactic grammars always benefit from access to complex segment representations. A learner of English phonotactics does better when it sees [tʃ, dʒ] as affricates than as clusters. But it does better still with [ts, dz] as affricates, and with prenasalized and labiovelar stops (*ju<u>mb</u>o, rug<u>b</u>y*).

Devising a phonotactics-based alternative requires nontrivial decisions about structuring the hypothesis space. The learner might need to entertain many types of complex segments, whose inventory could be gapped. For the English learner, what order should the learner use for [tʃ, dʒ, ts, dz]? What about possible three- and four-part complex segments (e.g. [ntʃw, ndʒw])? We set those questions aside, and tested the phonotactics-based strategy by manually creating progressively more elaborate representations, and training the UCLA Phonotactic Learner (UCLAPL, Hayes & Wilson 2008 et seq.)[19] on the resulting datasets. We evaluated the phonotactic grammars' fit with two measures: (i) the **log probability** of the data; (ii) **generality**: average number of segments covered by each constraint. Log probability is calculated by the learner for the entire grammar after each constraint is added (Hayes & Wilson 2008:386–387); we use the final, highest value. Generality seems to us to characterize good phonological grammars.

We tested this approach on phonotactically restrictive languages (Fijian, Ngbaka, Mbay) and permissive ones (Russian, English). In all the languages, replacing clusters with linguist-posited complex segments improves phonotactic grammar fit. The trouble is that still more improvement resulted when we added segments not usually posited for those languages.

The English results are in Table 24. The first simulation assumes only singletons, transcribing *angel* as [eɪ n͡ d͡ ʒ ə l]; it would be [eɪ n͡ dʒ ə l] in (d), and [eɪ ndʒ ə l] in (f). Simulation results are arranged in order of improved fit, showing that fit improves with the number of complex segments assumed. According to this implementation of the phonotactic strategy, English has prenasalized stops and affricates.

	Complex segments	No. of constraints	Generality	Log probability
a.	—	64	7.21	1,341,862
b.	[tʃ, dʒ]	53	7.40	1,374,595
c.	[ts, dz]	53	7.40	1,355,595
d.	[mb, nd, ŋg, mp, nt, ŋk]	60	7.82	1,379,973
e.	[ts, dz, tʃ, dʒ]	52	8.58	1,382,844
f.	[mb, nd, ŋg, mp, nt, ŋk, ts dz, tʃ, dʒ, ntʃ, nts, ndʒ, ndz]	59	10.14	1,413,231

Table 24: English phonotactic simulations assuming different complex segment inventories

This bizarre result arises in part[20] because rewriting clusters as segments allows the learner to capture gaps in a general way. For the phonotactically restrictive Fijian, the (C)V analysis relies on the inventory including [mb, nd, ŋg, tʃ, ndʒ, nr]. If the learner sees these sequences as segments, it can posit a general constraint *[-syll][-syll] to capture the absence of clusters. If the learner sees the sequences as clusters, however, it must induce many more constraints to explain which clusters are missing: *[+cont][-syll], *[-nas][-cont], *[+nas][-voice], etc.

Similar logic extends to more phonotactically permissive languages such as English. English has more homorganic NCs than heterorganic ones. If homorganic NCs are prenasalized stops, the learner can explain why heterorganic NCs are rare with *[+nas][-son]. If all NCs are clusters, the

[19] We used the gain version, not the Observed/Expected version; see Gouskova & Gallagher (2020).

[20] Another reason is a design feature of the UCLAPL: it does better on shorter words (Daland 2015). Representing clusters as complex segments reduces average word length and improves model fit, even at the cost of the increased number of natural classes the learner must navigate.

learner needs constraints for each type of heterorganic NC (cf. Wilson & Gallagher 2018). And because the UCLAPL searches unigram constraints before bigrams or trigrams, treating certain consonant sequences as segments makes it easier for the learner to discover constraints that hold over those sequences. The English grammar in Table 24f includes a constraint against prenasalized affricates, since those are rare in the language. The same pattern held in Russian. This is a perverse outcome: the more restricted the complex segment, the better the fit. Thus, the phonotactic strategy supplies the biggest motivation for adding complex segments in cases where the distributional evidence for complex segments is lacking. This casts a serious doubt on phonotactics as a learning theory of complex segments.

6.4 Learning complexity from phonetics

We now consider the hypothesis that complex segments differ from clusters phonetically. Trubetzkoy (1939) first conjectured that clusters are longer than complex segments, and much subsequent research has looked for duration differences. Within a learning theory, duration differences could give the learner a clue: unify short sequences but not long ones. As we will show, however, the existing phonetic research into duration differences raises more questions than it answers.

We discuss two reasons to doubt a universal correlation between segmenthood and duration. First, there are clear counterexamples. Second, segments and clusters can differ quite drastically in inherent duration, both within and across languages. There is often no principled way to decide what durations to compare. We end with a brief discussion of additional phonetic properties that could potentially differentiate segments from clusters.

Counterexamples. There are some reported correlations between duration and segmenthood (Brooks 1964, Riehl 2008, Cohn & Riehl 2016), but there are also counterexamples. We discuss two. First is Javanese (Adisasmito-Smith 2004), where NCs are longer than single segments but appear to have the distribution of segments. Second is Bura, where the same contradiction appears for labiovelars. The discussion of Javanese follows Stanton 2017:57–59. The Bura discussion is based on Maddieson 1983 and Sagey 1986:180–184.

Evidence from phonotactics and alternations in Javanese suggests that NCs pattern as single segments. NCs are the only initial clusters (though they result from prefixation; see Adisasmito-Smith 2004:258). NC clusters can combine with liquids medially, just like single stops. NCs also condition vowel reduction like singletons: as shown in 13, [i, u, ɔ] appear in open syllables, and [ɪ, ʊ, a] in closed ones (cf. [pʱʊ̲k.ti̲] and [pʱu̲.kɪ̲t]). NC sequences, shown in 14, are preceded by [i, u], just like [t, k] in 13 and unlike [k t, r n].

(13) Vowel centralization in closed syllables (Adisasmito-Smith 2004:261)

a.	[titɪp]	‘leg’	cf.	a’.	[titi]	‘meticulous’
b.	[kukʊr]	‘scratch’		b’.	[kuku]	‘finger’
c.	[pʱʊkti]	‘evidence’		c’.	[pʱukɪt]	‘hill’
d.	[sɪrnɔ]	‘disappear’		d’.	[siram]	‘bathe’

(14) No vowel centralization before NC sequences (Adisasmito-Smith 2004:262–263)

a.	[tiŋkʱi]	‘louse’	c.	[liŋgʱɪs]	‘machete’
b.	[tuŋkʱu]	‘wait’	d.	[muŋkʊr]	‘face down’

The analytic arguments for treating Javanese NCs as complex segments are strong, and yet they are significantly longer than singleton stops and nasals in Javanese (Adisasmito-Smith 2004:307). Adisasmito-Smith (2004) ultimately claims that the evidence from distribution and alternations is misleading, and the sequences are represented as clusters. But the link between segmenthood and duration in Javanese is tenuous. The phonological arguments for segmenthood are straightforward, but unmatched by the duration patterns.

A second case of mismatch between duration and segmenthood comes from languages in the Bura-Margi cluster. These languages are claimed to have many complex segments, most controversial of which are the labiocoronals [pt, bd, mnpt, mnbd, ʔbd, pts, ptʃ] (Maddieson 1983:287). Contra segmental treatments (Hoffman 1963, Newman 1977), Maddieson (1983) argues that Bura labiocoronals are clusters on the basis of several phonetic criteria. First, they are sequentially articulated: the bilabial closure is released before the alveolar closure is complete. Second, 'the consonantal duration for /pt/ is considerably longer than the duration for a single /t/ or /p/' (Maddieson 1983:293). On this basis, he concludes that labiocoronals in Bura (and likely Margi) are clusters. But Sagey (1986:182–184) argues that distributionally, the labiocoronals pattern as single segments: they can appear as the second member of a medial cluster or the first member of an initial cluster, both places where sonority-violating clusters are otherwise illicit. Thus the Bura-Margi labiocoronals pattern like single segments, yet are longer than single segments.

These counterexamples cast doubt on a universal link between duration and segmenthood. Of course, one could claim that duration and not phonological evidence correctly diagnoses these sequences as clusters (Maddieson 1983, Adisasmito-Smith 2004, Riehl 2008). But this strikes us as circular: any research program linking segmenthood and duration needs independent evidence of segmenthood to use duration as a diagnostic. Clear evidence is hard to come by: first, because there is no field-wide consensus on criteria for distinguishing complex segments from clusters (Herbert 1986), and second, because the most straightforward evidence for duration/segmenthood link would have to come from languages that contrast complex segments with same-phone clusters (as in monomorphemic [t s i n]∼[ts i n]). If such contrasts exist, they are at best rare (Maddieson & Ladefoged 1993, Riehl 2008, and others).

Differences in inherent duration. Proponents of the duration diagnostic claim that complex segments are same duration as a single segment. But inherent durations vary both within and across languages. There are differences among segments. In English, fricatives are longer than stops and nasals, and sounds produced towards the front of the vocal tract are longer than those produced towards the back (Lehiste 1970, Umeda 1977). Nasals are considerably longer than stops in Sukuma (Maddieson & Ladefoged 1993:277) but not in English (Umeda 1977:848). There are also differences among clusters. Homorganic NC clusters are shorter than heterorganic ones in Dutch (Slis 1974) and several Australian languages (Stanton 2017:175–176). Homorganic [s t] is shorter than heterorganic [s p] and [s k] in Greek, but not in English (Arvaniti 2007:21–22). These differences suggest that there is no principled way to determine whether a sequence is a complex segment or a cluster by comparing it to similar sequences in other languages (see Riehl 2008:103–105 for discussion).

Inherent duration differences among the segments of a language also make it difficult to identify a principled reference point for "a single segment". Researchers who make such comparisons opt for different choices. Maddieson & Ladefoged (1993:270–271) compare the durations of prenasalized stops in Fijian to those of /t/, /k/, and /l/ (the "measurable intervocalic consonants"), whereas Riehl (2008:179) determines whether an NC is a segment or a cluster by comparing its

duration to a plain nasal at the same place of articulation. It is not obvious which approach is more principled. More generally, durational asymmetries among segment and cluster types raise the possibility that complex segments might be longer than simplex segments because they are just long segments. Likewise, true clusters might be shorter than some simplex segments due to cluster compression (Farnetani & Kori 1986). In short, there is no agreed-upon way to determine whether a sequence is a segment or a cluster by comparing its duration to simplex segments, nor is it clear that such a correlation would be meaningful in the first place.

Many of these points come up in work on Modern Greek [t s] and [d z]. The analysis of these sequences has been hotly debated, with evidence from phonotactics, morpho-phonology, and phonetics recruited in favor of opposing analyses (Joseph & Philippaki-Warburton 1987, Tzakosta & Vis 2007, Syrika et al. 2011; see especially Arvaniti 2007 for a review). The phonotactics of Greek do not provide conclusive evidence; there is no clear difference between [t s, d z] and other stop-fricative sequences. All can occur word-initially, and obey the same restriction on clustering (for example, none of [p s, k s, t s, d z] can precede a liquid). These patterns are consistent with either an affricate or a cluster analysis of [t s, d z]. The sequences have also been studied phonetically, with inconclusive results (see Arvaniti 2007 for critical discussion). By the duration diagnostic, we expect [t s] and [d z] to be shorter than [p s] and [k s], and indeed they are (Joseph & Lee 2010). As Arvaniti (2007) points out, however, this could be due to homorganicity: other studies demonstrate that uncontroversial clusters in Greek show the same asymmetry (i.e. [s t] is shorter than [s p, s k]). Arvaniti concludes that neither phonetics nor phonotactics provide clear evidence to adjudicate the status of the sequences.

We were therefore interested in testing our computational learner on Greek to see if the distributional evidence was any clearer. To test the learner, we transcribed an orthographic list of 59,325 lexemes from the Corpus of Modern Greek. The learner was unequivocal: [t s, d z] are clusters. We provide a partial table of inseparability measures in Table 25 (the learner identified 182 clusters). The two most inseparable sequences are [s t] (0.7) and [ŋ x] (0.44); the four stop-[s] clusters fall far below the threshold of 1 (other sequences with higher values are omitted for brevity). Thus, our results suggest that Greek [t s, d z] should be analyzed as clusters, not affricates.

	inseparability	N(C1C2)	N(C1)	N(C2)
s t	0.70	7832	74207	35284
ŋ x	0.44	114	167	5252
d z	0.23	275	4137	2377
k s	0.15	3292	28912	74207
p s	0.03	1234	20752	74207
t s	0.01	963	35284	74207

Table 25: Inseparability measures for Modern Greek

Other possible phonetic cues. For the reasons just enumerated, we do not believe that learners appeal to durational information to decide which consonant sequences are clusters and which are segments. We do not deny that durational information could be useful in individual cases,[21] but the lack of a clear correlation between duration and phonological patterning casts

[21]Brooks (1964) shows that duration is a cue to the cluster/affricate distinction in Polish [t̪ s̪] vs. [t̪s̪], for example.

doubt on duration as a universal diagnostic for segmenthood.

Other phonetic differences between complex segments and clusters are also not universal. Herbert (1986:134–139) observes that vowels often lengthen before NCs in languages where NCs are argued to be segments. But subsequent work on NCs has established that this apparent correlation between segmenthood and lengthening has exceptions. Vowels do not lengthen before Fijian prenasalized stops (Maddieson & Ladefoged 1993:272), and vowels are lengthened before nasal-stop clusters in Iraqw (Downing 2005). There is thus no clear correlation between length of a preceding vowel and the segmental status of an NC (Riehl 2008:108–112). Investigations of other correlates, such as the amount of nasalization in a preceding vowel, have also come up empty-handed (Riehl 2008:106–108).

One fundamental difference between our approach and phonetic investigations is that our approach works on a variety of complex segments. Conversely, no phonetic criteria (aside from duration) can consistently apply to all complex segments. One criterion is the lack of internal release (Jones 1918), but languages have different phonetic rules for releasing consonants in clusters (see Zsiga 2000 on English vs. Russian). Another criterion is simultaneous articulation, used as a diagnostic on labiovelar and labiocoronal sequences (Maddieson 1993, Zsiga & Tlale 1998, Chitoran 1998). But several types of complex segments—affricates and prenasalized stops—necessarily involve sequential articulation, so this criterion is useless for them.

It is of course possible that there are as-yet undiscovered phonetic properties that reliably distinguish complex segments from clusters, such as articulatory ones (Trubetzkoy's Rule II). Saltzman & Munhall 1989, Löfqvist 1991, Byrd 1996 and others hypothesize that a segment is a constellation of gestures with a stable timing pattern. Byrd (1996:160) suggests that this definition allows us to make predictions about differences between complex segments and clusters. For NCs, for example, the oral constriction and velum lowering gestures should be more stably coordinated in languages where they are prenasalized stops than in languages where they are clusters. Existing work has not identified consistent differences of this sort. On the one hand, Shaw and colleagues 2019 find timing differences between Russian [bʲ] and [br]. On the other hand, Browman and Goldstein 1986 look for articulation differences between Chaga [mb] and English [m b], and find none: 'both [. . .] are constellations involving a single bilabial closure gesture [. . .] How, then, given the similarity between their gestural structures, do we capture the distinction between prenasalised stops in Chaga and nasal-stop sequences in English? The simplest statement is as a distributional, or phonotactic, difference' (Browman & Goldstein 1986:235–236). We would contend that the distributional difference is inseparability.

7 Conclusion

To conclude, we set out to construct a theory of learning complex segments. We presented a computational learner that builds complex segments from distributional information, and illustrated its application to both language-internal and typological questions. On the typological front, we

The distinction is sometimes erroneously described as a same-place contrast (Clements & Keyser:35), but the [t] in the cluster is actually dental, whereas the fricative portion is retroflex. In the affricate, the entire sequence is retroflex (see Gouskova in preparation). For our learner, the two sequences would be completely distinct, such that [t̪s̪] is unified whereas [t̪ s̪] is not (something we have verified in a simulation on Polish; again, results in Gouskova (in preparation).

have shown that our learner can derive at least one generalization regarding the size of complex segments and suggested that it may help us explain other generalizations regarding their composition. On the language-internal front, our learner identifies complex segment inventories that align with phonological argumentation in most cases.

Trubetzkoy's original heuristics for deciding between complex segments vs. clusters considered separability into independent phones, inventory structure, phonetic differences, and phonotactic distributions. We argued that phonotactic distributions are easier to state once the learner has the right inventory, but they cannot be the basis of a learnability theory of complex segment representations. We also evaluated the evidence for phonetic differences between complex segments and clusters, and concluded that these differences are not consistent enough to be a learning cue. We believe our model supplies an objective test to adjudicate between complex segments and clusters in languages where the evidence from other heuristics is inconclusive.

The next steps in this long-standing line of phonological research must involve securing better behavioral or psycholinguistic evidence for mental representations. In languages such as Quechua, the phonotactic patterning is clear and has been probed experimentally, with interpretable results. But for the vast majority of languages where complex segments are posited, good experimental evidence is missing. We are also far from an understanding of exactly what data are used in phonological learning. The dictionary-like lists that have become the norm are convenient but not obviously right as a model of learning data. Only when we have analytic, computational, and experimental results converging on the same conclusions, can we be sure that the hypothesized phonological representations are real.

References

Adisasmito-Smith, Niken. 2004. *Phonetic and Phonological Influences of Javanese on Indonesian.* Ithaca, NY: Cornell University dissertation.

Adriaans, Frans, and René Kager. 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language* 62.311–331.

Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26.9–41.

Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90.119–161.

Arvaniti, Amalia. 2007. Greek phonetics: The state of the art. *Journal of Greek linguistics* 8.97–208.

Asherov, Daniel, and Outi Bat-El. 2019. Syllable structure and complex onsets in Modern Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics* 11.69–95.

Asherov, Daniel, and Evan-Gary Cohen. 2019. A phonetic description of Modern Hebrew consonants and vowels. *Brill's Journal of Afroasiatic Languages and Linguistics* 11.3–27.

Baayen, R Harald; Richard Piepenbrock; and H Rijn van. 1993. The {CELEX} lexical data base. {CD-ROM}.

Barlow, Jessica A. 2001. Individual differences in the production of initial consonant sequences in pig latin. *Lingua* 111.667–696.

Becker, Michael, and Maria Gouskova. 2016. Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry* 47.391–425.

Béland, Renée, and Régine Kolinsky. 2005. One sound heard as two: The perception of affricates in Quebec French by Belgian French speakers. *Journal of Multilingual Communication Disorders* 3.110–117.

Blust, Robert. 1997. Nasals and nasalization in Borneo. *Oceanic Linguistics* 36.149–179.

Bolozky, Shmuel. 1980. On the monophonematic interpretation of Modern Hebrew affricates. *Linguistic Inquiry* 11.793–799.

Bolozky, Shmuel, and Michael Becker. 2006. Living lexicon of Hebrew nouns. Online resource. Available at http://sublexical.phonologist.org/.

Borowsky, Toni. 1986. *Topics in the Lexical Phonology of English.* University of Massachusetts, Amherst dissertation.

Brooks, Maria Zagórska. 1964. On Polish Affricates. *Word* 20.207–210.

Browman, Catherine P., and Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6.201–251.

Browman, Catherine P., and Louis M. Goldstein. 1986. Towards an articulatory phonology. *Phonology* 3.219–252.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10.425–455.

Byrd, Dani. 1996. A phase window framework for articulatory timing. *Phonology* 13.139–169.

Chimhundu, H. 1996. *Duramazwi reChiShona.* Harare: College Press Publishing Ltd. Online: `https://www.dokpro.uio.no/allex/gsd.html`.

Chitoran, Ioana. 1998. Georgian harmonic clusters: Phonetic cues to phonological representation. *Phonology* 15.121–141.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English.* New York: Harper & Row.

Clements, George N., and Samuel Jay Keyser. *CV phonology: A generative theory of the Syllable.* Cambridge, MA: MIT Press.

Cohn, Abigail. 1992. The consequences of dissimilation in Sundanese. *Phonology* 9.199–220.

Cohn, Abigail C., and Anastasia K. Riehl. 2016. Are there post-stopped nasals in Austronesian? *Studies in language typology and change*, ed. by Yanti and Timothy McKinnon, *NUSA*, vol. 60, 29–57. Tokyo: Tokyo University of Foreign Studies.

Cover, Thomas M, and Joy A Thomas. 2012. *Elements of information theory.* Hoboken, NJ: John Wiley & Sons.

Cowan, N. 1989. Acquisition of Pig Latin: A case study. *Journal of Child Language* 16.365–386.

Daland, Robert. 2015. Long words in maximum entropy phonotactic grammars. *Phonology* 32.353–383.

Daland, Robert; Bruce Hayes; James White; Marc Garellek; Andrea Davis; and Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28.197–234.

Danis, Nicholas Stephen. 2017. *Complex place and place identity.* New Brunswick, NJ: Rutgers University dissertation.

Devine, A. M., and Laurence D. Stephens. 1977. *Two studies in Latin Phonology.* Saratoga, CA: Anma Libri and Co.

Dixon, Robert M. W. 1981. Wargamay. *Handbook of Australian Languages*, ed. by Robert M. W.

Dixon and B. J. Blake, vol. 2. Amsterdam: John Benjamins.
Dixon, Robert M. W. 1988. *A Grammar of Boumaa Fijian.* Chicago: University of Chicago Press.
Doke, Clement Martyn. 1931. *A comparative study in Shona phonetics.* University of the Witwatersrand Press.
Downing, Laura J. 2005. On the ambiguous segmental status of nasals in homorganic NC sequences. *The internal organization of phonological segments*, 183–216.
Farnetani, Edda, and Shiro Kori. 1986. Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication* 5.17–34.
Fortune, George. 1980. *Shona Grammatical Constructions Vol. I.* 2 edn. Harare: Mercury Press.
Fougeron, Cécile, and Caroline L Smith. 1993. French. *Journal of the International Phonetic Association* 23.73–76.
Frisch, Stefan A.; Janet B. Pierrehumbert; and Michael B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language & Linguistic Theory* 22.179–228.
Gallagher, Gillian. 2011. Acoustic and articulatory features in phonology–the case for [long VOT]. *The Linguistic Review* 28.281–313.
Gallagher, Gillian. 2013. Learning the identity effect as an artificial language: Bias and generalisation. *Phonology* 30.253–295.
Gallagher, Gillian. 2016. Asymmetries in the representation of categorical phonotactics. *Language* 92.557–590.
Gallagher, Gillian. 2019. Phonotactic knowledge and phonetically unnatural classes: the plain uvular in cochabamba quechua. *Phonology* 36.37–60.
Gallagher, Gillian; Maria Gouskova; and Gladys Camacho-Rios. 2019. Phonotactic restrictions and morphology in Aymara. *Glossa* 4.1–39.
Garvin, Karee; Myriam Lapierre; and Sharon Inkelas. 2018. A Q-Theoretic approach to distinctive subsegmental timing. *Proceedings of the Linguistic Society of America* 3.1–13.
Gelman, Susan A; Bruce Mannheim; Carmen Escalante; and Ingrid Sanchez Tapia. 2015. Teleological talk in parent–child conversations in Quechua. *First Language* 35.359–376.
Göksel, Asli, and Celia Kerslake. 2004. *Turkish: A comprehensive grammar.* New York: Routledge.
Gordon, Matthew K. 2016. *Phonological Typology.* No. 1 in Oxford Surveys in Phonology and Phonetics. Oxford: Oxford University Press.
Gouskova, Maria. 2004. Relational hierarchies in Optimality Theory: The case of syllable contact. *Phonology* 21.201–250.
Gouskova, Maria. 2012. Unexceptional segments. *Natural Language and Linguistic Theory* 30.79–133.
Gouskova, Maria, and Michael Becker. 2013. Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language & Linguistic Theory* 31.735–765.
Gouskova, Maria, and Gillian Gallagher. 2020. Inducing nonlocal constraints from baseline phonotactics. *Natural Language & Linguistic Theory* 38.77–116. Online: `https://doi.org/10.1007/s11049-019-09446-x`.
Gravina, Richard. 2014. *The phonology of Proto-Central Chadic: the reconstruction of the phonology and lexicon of Proto-Central Chadic, and the linguistic history of the Central Chadic languages.* Utrecht: LOT.
Halle, Morris. 1959. *The Sound Pattern of Russian.* The Hague: Mouton.

Halle, Morris, and Alec Marantz. 1993. Distributed Morphology. *The View from Building 20. Essays in Honor of Sylvain Bromberger*, ed. by Kenneth Hale and Samuel Jay Keyser. Cambridge, MA: MIT Press.

Hayes, Bruce, and Tanya Stivers. 1996. The phonetics of post-nasal voicing. Ms., UCLA.

Hayes, Bruce, and James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44.45–75.

Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* 39.379–440.

Heath, Teresa. 2003. Makaa (A83). *The Bantu Languages*, ed. by Derek Nurse and Gérard Philippson, 335–348. Abingdon: Routledge.

Henrix, Marcel. 2015. *Dictionnaire Ngbaka-Français*. Munich: Lincom GmbH.

Herbert, R. K. 1986. *Language Universals, Markedness Theory, and Natural Phonetic Processes*. New York: Mouton de Gruyter.

Hester, Elizabeth, and Barbara Williams Hodson. 2004. The role of phonological representation in decoding skills of young readers. *Child language teaching and therapy* 20.115–133.

Hoffman, Carl. 1963. *A Grammar of the Margi Language*. London: Oxrford University Press.

Inkelas, Sharon; Aylin Küntay; Orhan Orgun; and Ronald Sprouse. 2000. Turkish electronic living lexicon (TELL). *Turkic Languages* 4.253–275.

Inkelas, Sharon, and Stephanie Shih. 2016. Re-representing phonology: consequences of Q Theory. *NELS 46: Proceedings of the Forty-Sixth Annual Meeting of the North East Linguistic Society*, ed. by Christopher Hammerly and Brandon Prickett, 161–174. Amherst, MA: GLSA.

Inkelas, Sharon, and Stephanie S Shih. 2013. ABC+ Q: Contour segments and tones in (sub) segmental Agreement by Correspondence. *21st Manchester Phonology Meeting*.

Ito, Junko. 1986. *Syllable Theory in Prosodic Phonology*. University of Massachusetts, Amherst dissertation. Published 1988. Outstanding Dissertations in Linguistics series. New York: Garland.

Jakobson, Roman. 1929. Remarques sur l'evolution phonologique du russe comparee a celle des autres langues slaves. *Travaux du Cercle linguistique de Prague* 2, Reprinted in Selected Writings of Roman Jakobson, Vol. 1.

Jakobson, Roman; Gunnar Fant; and Morris Halle. 1952. *Preliminaries to Speech Analysis*. Cambridge, MA: MIT Press.

Jones, Daniel. 1918. *An Outline of English Phonetics*. Cambridge: W.Heffer & Sons.

Joseph, Brian D, and Gina M Lee. 2010. Greek ts/dz as internally complex segments: Phonological and phonetic evidence. *Ohio State University Working Papers in Lingiustics*, ed. by Marivic Lesho, Bridget J. Smith, Kathryn Campbell-Kibler, and Peter Culicover, vol. 59, 1–10. Columbus, OH: Ohio State University. Department of Linguistics.

Joseph, Brian D, and Irene Philippaki-Warburton. 1987. *Modern Greek*. Croom Helm London.

Kadenge, Maxwell. 2010. Complexity in phonology: The complex consonants of simple CV-syllables in Zezuru. *Southern African Linguistics and Applied Language Studies* 28.393–408.

Kahn, Daniel. 1976. *Syllable-based Generalizations in English Phonology*. Cambridge, MA: MIT dissertation. Published by Garland Press, New York, 1980.

Kaplan, Abby. 2007. If *NT and *ND got into a fight, who would win? Ranking paradoxes and English postnasal stop deletion. *Phonology at Santa Cruz*, ed. by Aaron F. Kaplan and David Teeple, vol. 7, 25–35. Research Center, UC Santa Cruz.

Keegan, John M. 1996. *Dictionary of Mbay*. Lincom Europa.

Keegan, John M. 1997. *A reference grammar of Mbay*, vol. 14. Lincom Europa.

Kehrein, Wolfgang. 2013. *Phonological representation and phonetic phasing: affricates and laryngeals*, vol. 466. Walter de Gruyter.

Kochetov, Alexei. 2006. Syllable position effects and gestural organization: Articulatory evidence from Russian. *Laboratory Phonology 8*, ed. by Louis Goldstein, D. H. Whalen, and Catherine T. Best, 565–588. Mouton De Gruyter.

Kornfilt, Jaklin. 2013. *Turkish*. Routledge.

Krämer, Martin. 2007. *The phonology of Italian*. Oxford: Oxford University Press.

Ladefoged, Peter. 1996. *Elements of Acoustic Phonetics*. 2 edn. Chicago: University of Chicago Press.

Ladefoged, Peter, and Ian Maddieson. 1996. *The sounds of the world's languages*, vol. 1012. Oxford: Blackwell.

Laime Ajacopa, Teofilo. 2007. *Diccionario bilingüe; Iskay simipi yuyayk'ancha Quechua–Castellano, Castellano–Quechua*. La Paz, Bolivia.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.

Lembaga Basa, and Sastra Sunda. 1985. *Kamus Umum Basa Sunda*. Indonesia: Penerbit Tarate Bandung.

Lewis, Charlton Thomas; William Freund; and Charles Short. 1969. *A latin dictionary: founded on andrews' edition of freund's latin dictionary*. Clarendon Press.

Lin, Yen-Hwei. 2011. Affricates. *Blackwell companion to phonology*, ed. by Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice, vol. I, 367–390. Wiley-Blackwell.

Löfqvist, Anders. 1991. Proportional timing in speech motor control. *Journal of Phonetics* 19.343–350.

Lombardi, Linda. 1990. The nonlinear organization of the affricate. *Natural Language & Linguistic Theory* 8.375–425.

MacEachern, Margaret. 1997. *Laryngeal Cooccurrence Restrictions*. Los Angeles: UCLA dissertation.

Maddieson, Ian. 1983. The analysis of complex phonetic elements in Bura and the syllable. *Studies in African Linguistics* 14.285–310.

Maddieson, Ian. 1990. Shona velarization: complex consonants or complex onsets? *UCLA Working Papers in Linguistics* 74.16–34.

Maddieson, Ian. 1993. The Structure of Segment Sequences. *UCLA Working Papers in Phonetics*, vol. 83, 1–7. UCLA Department of Linguistics.

Maddieson, Ian, and Peter Ladefoged. 1993. Partially nasal consonants. *Nasals, Nasalization, and the Velum*, ed. by Marie Huffman and Rena Krakow, 251–301. San Diego, CA: Academic Press.

Maes, Védaste. 1959. *Dictionnaire ngbaka-français-néerlandais: précédé d'un aperçu grammatical*. Tervuren: Musée royale de Congo belge.

Martinet, André. 1939. Un ou deux phonèmes? *Acta linguistica* 1.94–103.

McCrary, Kristie Marie. 2004. *Reassessing the Role of the Syllable in Italian Phonology: An Experimental Study of Consonant Cluster Syllabification, Definite Article Allomorphy and Segment Duration*. Los Angeles, CA: UCLA dissertation.

McCullagh, Matthew. 2011. The Sounds of Latin: Phonology. *A Companion to the Latin Lan-*

guage, ed. by James Clackson, 81–91. Hoboken, NJ: Blackwell Publishing Ltd.

Mohanan, K. P. 1982. *Lexical Phonology*. Cambridge, MA: MIT dissertation. Distributed by IULC Publications.

Moreton, Elliott. 2002. Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84.55–71.

Mudzingwa, Calisto. 2010. *Shona morphophonemics: Repair strategies in Karanga and Zezuru*. Vancouver, BC: University of British Columbia Doctoral dissertation.

Newman, P. 1977. Chadic classification and reconstruction. *Afroasiatic Linguistics* 5.1–42.

Newman, Paul, and Roxana Ma. 1966. Comparative Chadic: phonology and lexicon. *Journal of African Languages and Linguistics* 5.218–251.

Nurse, Derek, and Gérard Philippson. 2006. *The Bantu languages*. New York: Routledge.

Ohala, John J., and James Lorentz. 1977. The story of [w]: an exercise in the phonetic explanation for sound patterns. *Proceedings of the 3rd Annual Meeting of the Berkeley Linguistics Society*, ed. by Kenneth Whistler, Robert jr. van Valin, Chris Chiarello, Jeri J. Jaeger, Miriam Petruck, Henry Thompson, Ronya Javkin, Anthony Woodbury, Kenneth Whistler, Robert jr. van Valin, Chris Chiarello, Jeri J. Jaeger, Miriam Petruck, Henry Thompson, Ronya Javkin, and Anthony Woodbury, 577–599. Berkeley: Berkeley Linguistics Society.

Padgett, Jaye. 1995a. Partial class behavior and nasal place assimilation. *Proceedings of the Arizona Phonology Conference: Workshop on Features in Optimality Theory*. Available on Rutgers Optimality Archive, http://roa.rutgers.edu.

Padgett, Jaye. 1995b. *Stricture in Feature Geometry*. Stanford: CSLI Publications.

Padgett, Jaye. 2003. Contrast and postvelar fronting in Russian. *Natural Language & Linguistic Theory* 21.39–87.

Padgett, Jaye, and Marzena Żygis. 2007. The evolution of sibilants in Polish and Russian. *Journal of Slavic linguistics*, 291–324.

Parker, Steve, and David Weber. 1996. Glottalized and aspirated stops in Cuzco Quechua. *International Journal of American Linguistics* 62.70–85.

Pater, Joe. 1999. Austronesian nasal substitution and other NC effects. *The Prosody-Morphology Interface*, ed. by René Kager, Harry van der Hulst, and Wim Zonneveld, 310–343. Cambridge: Cambridge University Press.

Phillips, Lawrence, and Lisa Pearl. 2015. The utility of cognitive plausibility in language acquisition modeling: Evidence from word segmentation. *Cognitive science* 39.1824–1854.

Pope, Mildred Katharine. 1934. *From latin to modern french with especial consideration of anglo-norman: Phonology and morphology*. 6. Manchester University Press.

Riehl, Anastasia Kay. 2008. *The phonology and phonetics of nasal obstruent sequences*. Ithaca, NY Doctoral dissertation.

Robins, R. H. 1959. Nominal and verbal derivation in Sundanese. *Lingua* 8.337–369.

Robins, R.H. 1957. Vowel nasality in Sundanese: A phonological and grammatical study. *Studies in Linguistic Analysis*, ed. by J.R. Firth, 87–103. Oxford: Blackwell.

Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80.475–532.

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

Sagey, Elizabeth. 1986. *The Representation of Features and Relations in Nonlinear Phonology*.

Cambridge, MA: MIT dissertation. Published by Garland Press, New York, 1991.

Saltzman, Elliot L., and Kevin G. Munhall. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1.333–382.

Scholes, Robert J. 1966. *Phonotactic grammaticality.* Berlin: Mouton.

Schwarz, Martha; Myriam Lapierre; Karee Garvin; and Sharon Inkelas. 2019. Recent advances in Q Theory: Segment strength. *Proceedings of the Linguistic Society of America* 4.1–14.

Selkirk, Elisabeth. 1982. The syllable. *The Structure of Phonological Representations*, ed. by Harry van der Hulst and Norval Smith, vol. Part II. Dordrecht: Foris Publications.

Shaw, Jason A; Karthik Durvasula; and Alexei Kochetov. 2019. The temporal basis of complex segments. *Proceedings of the 19th international congress of phonetic sciences*, ed. by Sasha Calhoun, Paola Escudero, Marija Tabain, and Paul Warren, Melbourne, Australia, 676–680.

Shih, Stephanie S, and Sharon Inkelas. 2019a. Autosegmental Aims in Surface-Optimizing Phonology. *Linguistic Inquiry* 50.137–196.

Shih, Stephanie S, and Sharon Inkelas. 2019b. Subsegments and the emergence of segments. *Proceedings of the Linguistic Society of America* 4.1–8.

Slis, I. H. 1974. Synthesis by Rule of Two-Consonant Clusters. *IPO Annual Progress Report*, ed. by A. J. Breimer, vol. 9, Eindhoven: Institute for Perception Research, 64–69.

Stanton, Juliet. 2017. *Constraints on the distribution of nasal-stop sequences: An argument for contrast.* Massachusetts Institute of Technology dissertation.

Stanton, Juliet. 2019. Allomorph selection precedes phonology: evidence from Yindjibarndi. Talk presented at the 27th Manchester Phonology Meeting.

Steriade, Donca. 1993. Closure, release, and other nasal contours. *Nasals, nasalization, and the velum*, ed. by Marie K. Huffman and Rena A. Krakow, 401–470. San Diego: Academic Press.

Steriade, Donca. 2012. Intervals vs. syllables as units of linguistic rhythm. Handouts, EALING, Paris.

Syrika, Asimina; Katerina Nicolaidis; Jan Edwards; and Mary E Beckman. 2011. Acquisition of initial/s/-stop and stop-/s/sequences in Greek. *Language and speech* 54.361–386.

Thomas, Jacqueline M. C. 1963. *Le parler ngbaka de Bokanga.* Paris: Mouton.

Tikhonov, A. N. 2002. *Morfemno-orfografičeskij slovar.* Moscow: ACT Publishing.

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

Tlale, One. 2005. *The phonetics and phonology of Sengwato, a dialect of Setswana.* Washington, D.C.: Georgetown University dissertation.

Townsend, Charles, and Laura Janda. 1996. *Common and Comparative Slavic: Phonology and Inflection.* Columbus: Slavica Publishers.

Trubetzkoy, N. S. 1939. *Grundzüge der Phonologie.* Prague: Travaux du cercle linguistique de Prague 7.

Tzakosta, Marina, and Jeroen Vis. 2007. Phonological representations of consonant sequences: the case of affricates vs."true" clusters. Paper presented at the 8th Iinternational Conference on Greek linguistics, Ioannina, Greece.

Umeda, Noriko. 1977. Consonant duration in American English. *The Journal of the Acoustical Society of America* 61.846–858.

Van De Weijer, Jeroen. 2011. Secondary and double articulation. *The Blackwell companion to phonology*, 694–710.

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin.* Berlin: Mouton de Gruyter.

Wilson, Colin, and Gillian Gallagher. 2018. Accidental gaps and surface-based phonotactic learning: A case study of South Bolivian Quechua. *Linguistic Inquiry* 49.610–623.

Wordick, Frank. 1982. *The Yindjibarndi Language.* Canberra: Australian National University.

Zaliznjak, Andrej Anatoljevich. 1977. *Grammatičeskij slovar' russkogo jazyka. [A grammatical dictionary of the Russian language].* Moscow: Russkij Jazyk.

Zsiga, Elizabeth. 2000. Phonetic alignment constraints: consonant overlap in English and Russian. *Journal of Phonetics* 28.69–102.

Zsiga, Elizabeth, and One Tlale. 1998. Labio-coronal fricatives in Sengwato. *Journal of the Acoustical Society of America* 104.1779.

Zuraw, Kie. 2000. *Patterned exceptions in phonology.* Los Angeles, CA: UCLA dissertation.