Learning Complex Segments*

Maria Gouskova & Juliet Stanton
New York University

May 27, 2020

Abstract

Languages differ in the status of sequences such as [mb, kp, ts]: they can pattern as complex segments or as clusters of simple consonants. We ask what evidence learners use to figure out which representations their languages motivate. We present an implemented computational model that starts with simple consonants only, and builds more complex representations by tracking statistical distributions of consonant sequences. We demonstrate that this strategy is successful 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, tʃʰ, tʃ'] pattern as single segments—affricates, part of the natural class of stops. First, affricates pattern with stops in 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 to these generalizations. Second, the complex segment analysis is justified by the phonemic inventory of Quechua: affricates show a three-way laryngeal contrast just like other stops, [p p' pʰ t t' tʰ k k' kʰ]. The fricatives [ʃ, s, x], on the other hand, do not have laryngeal contrasts—[ʃ'] and [ʃʰ] do not occur except in affricates. Finally, affricates participate in non-local cooccurrence restrictions on ejectives and aspirates, which cannot follow other stops within the same phonological word (MacEachern 1997, a.o.). None of these generalizations could be stated elegantly without assuming complex segments.

*Please address correspondence to both authors. We would like to thank Michael Becker, Lisa Davidson, Gillian Gallagher, and Maddie Gilbert, and Donca Steriade for comments on the manuscript. We are also grateful to Adam Albright for Latin paradigms, Michael Becker for Turkish materials (and the suggestion to test the learner on Hebrew), 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, in brief (data from Gallagher 2016a, 2019; Gouskova and Gallagher 2020)

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

misk'i	'delicious'	*miksi		but cf.	k'atʃa	'pretty'
tʃ'impu	'boil'	*tsimpu, *tintsu		but cf.	antʃa	'a lot of'

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

tʃ'uspi	'to fly'	*tʃ'usp'i	*p'uk'i
satʃ'a	'tree'	*patʃ'a	*kap'a
kʰutʃi	'pig'	*kʰutʃʰi	*kʰutʰi

By contrast, European French has only simplex segments (Fougeron and 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 the one we focus on in this paper. It is a non-trivial 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 most of 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, probabilistic 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. This measure 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 (see §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—i.e., *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 other 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 quite general.

The issue of the learning timeline is harder to address given the currently available evidence. We do not know when children acquire complex segment representations, and 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 (see §5.3, 6). 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 most of our simulations, we use dictionary-like lists of transcribed phonological words, where type frequencies of segments are represented (this is the approach taken in Albright and Hayes 2003; Hayes and Wilson 2008; Becker et al. 2011; Gouskova and Gallagher 2020). 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 §2.8, there are cases where results diverge significantly depending on the type of data the learner is given, and the morphemic lexicon emerges as the most appropriate model of the learning data.

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 and Inkelas 2018). Under our learnability-based proposal, this limitation follows naturally. If all complex segments result from unification, we would only expect long segments to be common if long clusters are, too. In reality, however, long clusters are rare both cross-linguistically and language-internally (see §4). As we show, when the learning data supply sufficient evidence of inseparability, as in Shona, our learner identifies complex segments with four parts. We argue that our proposal's ability to derive this 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 to the problem of learning complex segments: learning from phonotactics, and learning from phonetics (§5). Phonotactics seems initially promising: when the learner of, say, English is confronted with a sequence that could be a cluster or a segment (e.g., [ts], [tʃ] or [kp]), the learner tries phonotactic learning given a cluster representation and a complex segment representation, and assesses the fit of the resulting grammars. We argue that this approach is likely wrong, 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: the idea is that the learner uses some detectable differences in duration of clusters vs. complex segments, or perhaps whether a segment involves simultaneous or sequential articulations. We are skeptical of this alternative as a general strategy, as there are counterexamples where the phonetic evidence contradicts the phonotactics, and 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 a few terms. We assume that a *simplex segment* can be characterized using unique and non-conflicting feature specifications for place and manner: an [m] is nasal and labial through its articulation; a [t] involes a single 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] For example,

[1]Segments involving different laryngeal configurations (aspirates, ejectives, implosives) are sometimes treated as complex (Kehrein 2013; Shih and Inkelas 2018); we have no quarrel with this view, but do not treat them as sequences to be unified in our simulations, in order to keep the paper to a reasonable length.

[mb] is nasal for the initial part of its articulation but oral for the rest. Affricates such as [ts] involve a sequenced stop-like and fricative-like constriction. A labiovelar such as [kp] involves stop constrictions in more than one place of articulation, and thus is both labial (like [p]) and velar (like [k]). A labialized [kw] has a dorsal constriction with simultaneous lip rounding. The one unifying feature of complex segments is that they are articulatorily complex in a way that simplex segments are not, and they can be decomposed into parts that can be separate segments—either in the same language, or in other languages.

When we talk about clusters of simplex segments, we will henceforth refer to them as *clusters.* Complex segments are often written in a special way to highlight analytical assumptions or phonetic differences in duration: thus, the alveolar affricate can be [t͡s], [ʦ], or [t^{s}]; nasal-stop segments can be [n͡d], [^{n}d], or [n^{d}]. We eschew these orthographic conventions throughout—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. This distinction is due to Sagey (1986), for whom complex segments involve simultaneous articulation and contour segments are sequential. But this distinction is by no means universally accepted (see Lin 2011 for review). Thus, Lombardi (1990) argues that affricates are complex segments in Sagey's sense; their fricative and stop portions are unordered in the phonology, even though they are sequential in the phonetics. Even labiovelars such as [kp], which for Sagey would be complex segments, sometimes 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). More generally, the contour-complex distinction rests on analytic arguments as much as on phonetic reality. In labiovelars such as [kp], the labial and dorsal constrictions are near-simultaneous, but the releases of must be staggered in order for both places of articulation to be audible (Maddieson 1990; Ladefoged and Maddieson 1996). Constrictions can, of course, be near-simultaneous for clusters, as well (Browman and Goldstein 1989 show for English that in phrases like 'perfect memory', the [t] is inaudible because it is completely overlapped with surrounding stops). A labialized series [kw, tw, pw] is also articulatorily heterogeneous: simultaneous lip rounding and closure are impossible for [pw], but this does not have to affect its phonological patterning or analytical treatment. Moreover, it is possible for articulatorily and acoustically identical sequences to pattern as complex segments in one language and as clusters in another (Browman and Goldstein 1986; see also §5.3.1). Crucially, all of these sequences present the same problem for the learner.

We are open to the possibility that there are representational differences among what we call 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 of complex segments is detected, and concern ourselves only with this latter question.

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.

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* (§2.3), a *unification procedure* (§2.4), and *iteration* (§2.5). The algorithm is summarized as pseudocode in §2.6; §2.7 discusses the source of the sorts of distributions that allow the learner to be successful, and §2.8 considers the issue of the learning data.

2.1 The initial state

The learner is assumed to have two types of information available to it in the initial state. First, the learner has a lexicon of learning data containing only simplex segments. Figure 1 shows several representative examples of the initial state from the lexicon of Fijian. Our Fijian corpus is from the An Crúbadán project, http://crubadan.org (26,000 words, compiled from internet texts—we cleaned the corpus to exclude English words, which left us with 17,600).

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

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

The second type of information that the learner has in its initial state is a feature table of all simplex segments (access to features is a standard assumption in much work on computational phonology; see Albright and Hayes 2003; Hayes and Wilson 2008; Becker and Allen submitted; Gouskova and Gallagher 2020). For example, the learner knows that Fijian has [p], and that [p] is [-syllabic, -consonantal, +nasal, +sonorant, +voice, -continuant, -strident, +labial].

2.2 Background on Boumaa Fijian

The empirical focus of this section is on Fijian, so we give its consonant inventory in Table 1, following Dixon (1988:13).[2] Dixon posits six complex consonants for Fijian: /mb/, /nd/, /ŋg/, /nr/, /tʃ/, and /ndʒ/. The affricates /tʃ/ and /ndʒ/ appear mainly as allophonic variants of /t/ and /nd/ preceding /i/, but also occur outside of this context in several loanwords. Dixon's analysis of Fijian rests on a phonotactic argument. Aside from the complex segments, the language has no consonant sequences. Complex segments only occur prevocalically, just as singleton consonants. Under this analysis, Fijian has (C)V syllables.

2.3 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 (2) and C_2

[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	post-alveolar	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 segmental inventory (after Dixon 1988:13)

being in the cluster C_1C_2 (3). This calculation is essentially the same as Mutual Information (Cover and 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).[3] We discuss why the calculation must be over segments rather than sequences of natural classes in §5.1.

(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 major difference between Riehl's conception of inseparability and ours is that we treat inseparability as 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].

For a concrete example, 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].

[3]The limitation to [-syllabic] segments prevents our learner from positing complex vowel-consonant segments such as [op] and diphthongs such as [aʊ]. Since our focus is only on consonants, we leave to future work the question of whether the learner should be able to posit segments that disagree internally for [-syllabic] and [+syllabic].

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] (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 in the equation is the probaility of [ŋ g] (given all the clusters) divided by the probability of [g] (given all the consonants). This is roughly 8.317. (The value associated with [g] is higher than the value associated with [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] (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 in the equation is the probability of [n r] (given all the clusters) divided by the probability of [r] (given all the consonants). The product of these two terms is **0.91**.

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

Looking more broadly at the numbers in Table 3, there is a clear difference between [n r] and the rest of the clusters. This is because [b], [d], [g], [ʃ], and [ʒ] do not occur outside the listed 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.

2.4 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.[4] In order to qualify for unification, a sequence must satisfy two requirements:

- *Cluster inseparability must be equal to or greater than 1.* In order 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 α = .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 the actual frequency of a given cluster and all other clusters to the hypothetical frequencies were that cluster unattested. A sample contingency test for the least frequent cluster, [d ʒ], is in (7). 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).

(7) Sample Fisher's Exact Test for [d ʒ]

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

After the learner settles on a set of sequences to unify, it modifies the segmental inventory and its representation of the learning data. First, the learner modifies its feature table by adding the new complex segments and associated distinctive features.[5] Second, the learner modifies its lexicon by iteratively unifying eligible clusters, from most to least inseparable. This means that, for Fijian, the learner replaces [ŋ g] with [ŋg], then [m b] with [mb], and so on. Note that modifying the learning data in this way means that it is possible for unification of one cluster to bleed unification of another. In Fijian, for example, unification

[4]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 §5.2 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 §4 discusses ways to derive some of these same generalizations regarding the typology of complex segments.

[5]It is not obvious what the right featural representations for complex segments should be; indeed, most of the phonological work on complex segments concerns this problem (Anderson 1976; Rubach 1985; Lombardi 1990; Steriade 1993; Padgett 1995b; Clements and Hume 1995; Rubach 2000; Riehl 2008; Lin 2011 and many others). We acknowledge that a single strategy is unlikely to work for all languages, and that evidence from phonotactics and alternations must inform featural representations. But we had to make some practical decisions about reconciling features when they were in conflict between the two phones being unified. Sequencing both values would be the simplest solution, but we chose to pick one of the values, for practical reasons: this made our feature tables compatible with the SPE-style tables used in computational work such as Hayes and Wilson (2008). We represented affricates as [+strid,-cont], prenasalized stops as [+nas, -son], and gave labiovelars such as [kp] both [+LAB] and [+DOR] features. Secondary articulations such as labialization were represented as vocalic features on consonants. This approach conceptually separates the problem of which sequences should be unified from the problem of how to represent them featurally.

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]; all [n d ʒ]s are first converted to [nd ʒ] so there are no remaining [d ʒ]s to be replaced. Finally, the learner checks that each segment included 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.

2.5 Iteration

Following the unification procedure, the learner computes new frequencies for each segment and cluster, as well as inseparability measure for each cluster (according to the formula in §2.3). These values for the second iteration of the Fijian learning simulation are in Tables 4-5.

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 4: 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 5: Inseparability measures and CC counts for Fijian (second iteration)

Because there are only two clusters left ([nd ʒ] and [n r]), their inseparability measures are high. The total number of clusters has dropped, so the probability of the remaining clusters is much higher. (The sequence [nd ʒ] has a higher inseparability compared to [n r] because [ʒ] does not occur as a singleton, but [r] does.) The frequency 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 no longer exists in the lexicon). The third iteration finds no remaining clusters to convert to complex segments, and the learner converges on the inventory posited by Dixon (1988).

Iteration is necessary for two reasons, both of which are 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 (§4.3). 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 performs; it stops when it finds no more sequences that qualify for unification. It is thus capable of finding segments that contain three, four, five, or more subparts. In none of the cases we have investigated, however, does the learner actually find complex segments longer than four parts (see §4.3 on Shona). We discuss how this length-based restriction on complex segments is derived in §4.

2.6 Summary of the algorithm

The algorithm is given in pseudocode in (8). In prose form: 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 considered to be the best versions—no new 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.

(8) 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 by insep;

iv. If any insep(C1C2)≥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.

2.7 Why these distributions?

As we show in §3, our learner largely succeeds at finding the set of complex segments posited by linguists, often working from different criteria. But what causes complex segments to have these unique distributional properties that our learner picks up on?

We think this question likely has a multipart answer. First, it is likely that the learner succeeds in finding some complex segments from distributions alone because of their origins. From historical reconstructions and analyses of synchronic alternations, we know that some complex segments result from sound changes to singletons. Thus, affricates such as [ts] and [tʃ] are often the endpoint of palatalization of /t/ and /k/—as in Slavic (Jakobson 1929; Townsend and Janda 1996:76ff). Prenasalized stops can also have single-stop origins: for example, in modern Bantu, at least some prenasalized stops are reconstructed as deriving from Proto-Bantu voiced stops (Nurse and Philippson 2006:148). This hypothesis is supported by a comparison between Classical Latin (§3.2.1) and a pseudo-Italian corpus we created, where Latin /k, g, t, d/ became affricates before front vowels. When trained on the pseudo-Italian corpus, our learner readily identified all four resulting affricates: [tʃ, dʒ, ts, dz].

On the other hand, chains of historical changes can result in distributions that make discovering complex segments difficult. Thus, Latin stops became affricates in Old French and then became fricatives, with various other changes eliminating many original stops (Pope 1934:125ff). As a result, Modern European

French is not thought to have any affricates (§§1, 9). In a twist, Quebecois French reintroduced [ts, dz] as allophones of /t,d/ before high front vowels—and yet our computational learner does not consistently discover these affricates in all learning data, possibly because /t, d/ became too rare due to these chains of changes.

Sound change does not have to be the source of distributions like those of complex segments. In Muyang (Central Chadic), prenasalized [mb, nd, ndz, ŋg] coexist with voiced [b, d, dz, g] and nasal [m, n]. Prenasalized stops were innovated at some point between proto-Chadic (Newman and Ma 1966) and proto-Central Chadic (Gravina 2014), but it is not obvious 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 to a point where certain sequences can acquire the distributions of complex segments. Turkish borrowed [dʒ] from other languages so often that our learner readily identifies it as an affricate (§3.1.3). By contrast, Russian borrows [dʒ] as a cluster, and in our corpora, the sequence is far below the threshold for unification (§3.2.2). Thus, borrowing can but does not have to be a source of complex segments in a language.

2.8 On the nature of the learning data

Before diving into the remaining case studies, we consider two broader learnability questions: what data do learners actually use, and when does the learner arrive at complex segment representations? There is also a methodological issue of data quality, which haunts all computational modeling research. Our simulations for the most part rely on dictionary-like lists of phonological words—the easiest data for linguists to get, and the type widely used in statistical computational phonology (Zuraw 2000; Hayes and Wilson 2008, many others). But it is far from obvious that this is theoretically well-motivated.

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 thought to be relevant in morpho-phonological and phonotactic learning (Bybee 1995; Albright and Hayes 2003).[6] Correspondingly, phonotactic models such as the UCLA Phonotactic Learner are trained on a lexicon of phonological words (Hayes and Wilson 2008; Hayes and White 2013), much as in our Fijian example and many of the case studies below. On the other hand, Adriaans and Kager (2010) point out that a more realistic model would learn from connected speech: learners have to work out where words begin and end at the same time as learning phonotactics. Their model of speech segmentation uses token frequencies, on the assumption that morpheme/word concatenation creates sequences that are rarer inside words (e.g., [pk] is extremely rare morpheme-internally in English, as in "napkin", but allowed at morpheme/word boundaries, as in "upkeep" and "drip coffee"). There are alternatives to learning from whole words or unsegmented running speech: the SPE model of the lexicon contains morphemes, rather than whole words (Chomsky and Halle 1968). So, a model that uses an SPE-style lexicon would learn from morphemes, not word lists. For some kinds of morphologically sensitive phonotactics, learners might have to examine words but have access to morpheme boundaries (Gouskova and Gallagher 2020; Gallagher et al. 2018).

There are reasons to consider these alternatives seriously. As Adriaans and Kager point out, the "one word per line" lexicon is an idealization; it is itself the result of learning where word and morpheme bound-

[6] Bybee's original argument is morphological: highly irregular morphology, as in *go/went*, is supported by high token frequency, but regular morphology is type-frequent (it is what most verbs do). See fn. 10 of Hayes and Wilson 2008 for additional references.

aries are, and phonotactics aids segmentation. Research taking a broad view of phonological learnability must reconcile learning phonotactics, segmentation, and representations such as complex segments, which sometimes present a chicken-and-egg problem. As we will see in several case studies, some phonotactic constraints can be sensibly stated only when the learner has a good analysis of the segmental inventory of the language. So does the learner find complex segments first, or morpheme boundaries? Moreover, the distributions of complex segments sometimes can be characterized properly only with reference to morpheme boundaries. Is it possible that learners revise their segmental inventory after they become morphologically aware, much as has been argued for phonotactic grammars (Gouskova and Becker 2013 et seq.)?

Answering these questions would require systematic, in-depth investigation that relies on a variety of different kinds of evidence. We have only begun testing a variety of data for a handful languages: Quechua (roots, morphemes, phonological words; §3.1.4), Russian (pseudo-connected speech, phonological words, morphemes; §3.2.2), Quebecois French (phonological words vs. child-directed connected speech; §6), Hungarian and Navajo (phonological words vs. roots; see project website). With the exception of Quebecois French and Russian, all these languages point to the same conclusion: the more informative distributions are in morphemes/roots, not phonological words or connected speech. Quechua is the most striking example: because the language's phonotactics differ so much between affixes and roots, the frequency of certain clusters in morphologically complex words is so high that they end up incorrectly unified. When the learner is given roots or segmented, tokenized morphemes to learn from, it finds the posited inventory.

The morphological and phonotactic characteristics of languages do matter. Some languages' morphemes have about the same sound distributions as words, whereas others allow sequences inside morphemes that are not allowed at word edges. For example, Turkish morphemes have the same consonant phonotactics as Turkish words (Kornfilt 2013, p.495). The same appears to be true of Fijian (Dixon 1988, pp. 27–29). In other languages, such as Russian, morpheme concatenation is a significant source of sequences that could in principle be analyzed as complex segments (§3.2.2). These languages differ also in how easy it is to find morpheme boundaries and identify alternants of morphemes, presumably both for learners and for linguists and lexicographers. Morpheme concatenation can introduce sequences that are not found morpheme-internally, too: Mbay has prenasalized stops tautomorphemically, but tone-bearing nasals form clusters with stops at morpheme boundaries (see §3.1.2). English distinguishes affricates from non-homorganic sequences of stops and fricatives created by morpheme concatenation (see §3.2.3). Dealing with these subtleties requires a morphologically aware learning model—one that analyzes and unifies sequences in morphemes.

We want to end this discussion on an obvious methodological point: data quality matters. The better curated the corpus, the more likely the learner is to find complex segments that linguists posit. For example, we ran our learner on two versions of Navajo (Gallagher in prep.), which has a series of strident and lateral affricates. Our learner finds all nine of them [dʒ, tʃ', ts', tɬ', ts, tʃ, dɮ, dz, tɬ] when trained on a list of 917 stems, but it finds only [ts', dʒ] when trained on the 19,000-word-long An Crúbadán wordlist based on internet texts. Our simulations based on data prepared by lexicographers and linguists (Ngbaka, Shona, Russian, Turkish, Hebrew, etc.) tend to yield much cleaner results—which is unsurprising. But even clean, curated wordlists are not good models of the sort of data that a child encounters. Thus, in some cases, the sequences analyzed by linguists as complex segments are at the top of the queue for unification (e.g., [d z] and [t s] in Quebecois French, see §6), but do not the threshold for unification. Does this mean that Quebecois French provides ambiguous evidence as to the status of these sequences, or that our data source is of the wrong type, or that its quality is poor? In cases of this sort, we think examining the qualitative, not just the quantitative results is worthwhile. Whenever possible, therefore, we report results from multiple datasets, and we take qualitative patterns into account when we make conclusions on the basis of negative results (as in Latin, §3.2.1). But the most promising type of data—morpheme lists—is also the hardest to create, so we have to leave the question of learning data ultimately open.

3 Case studies

In addition to Fijian, we tested our learner on a large range of languages (twenty-three as of this writing). We separate the cases into three analytic groups. The first group includes languages where the phonotactic arguments for complex segments are clear: Ngbaka, Mbay, Turkish, Hebrew, and Quechua (§3.1). For these languages, our learner arrives at the segmental inventories that align with the analytic arguments. The second group includes languages in which complex segments have been posited despite a lack of clear arguments. We discuss three such cases (Latin, Russian, English) in §3.2. Unsurprisingly, our learner's findings for these languages are more mixed; it finds complex segments in some but not others, and the results are more fragile. The third type of case is represented by Sundanese (§3.3). In Sundanese, analysts posit a different set of complex segments from what our learner finds. Additional case studies are discussed in §4–6, in the context of typological, phonetic, and experimental considerations.

Before continuing, a note is necessary regarding nature of our claims, and how these relate to the type of information available to us (namely, descriptive grammars and articles). Authors of phonological descriptions are often not explicit about their reasons for assuming a complex segment or a cluster analysis, leaving it up to the reader to determine if their conclusions are warranted. Given a lack of clear criteria, this is difficult to do. Our claim in this section is not that our learner correctly diagnoses the presence of complex segments in any given language, because linguists' descriptions of these representations are not necessarily correct. Rather, our claim is that our learner returns results that are compatible with linguists' descriptions. Given that linguists often reason on the basis of very different metrics than those our learner uses, we think this result is striking. Issues inherent in accessing speaker knowledge more directly are discussed in §6.

3.1 Confirming the complex segment analysis

In both Ngbaka (§3.1.1) and Mbay (§3.1.2), arguments for segmenthood are straightforward: complex segments have the distribution of equivalent simplex segments and not of clusters. Unlike Fijian, however, all complex segments in Ngbaka and Mbay are fully separable, and both languages allow true clusters (though to different extents). Despite these additional difficulties, our learner successfully acquires the target inventories for both languages. Next are two brief case studies from Turkish and Hebrew (§3.1.3). We end with Quechua, a language that offers multiple arguments for the complex segment analysis but that also demonstrates that the results depend on the kind of data that the learner is exposed to.

3.1.1 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 (9)). The consonant inventory of Ngbaka is in Table 6.[7]

One argument for complex segments is phonotactic. They are both reasonably common and allowed in all places where 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 in a way that would be difficult to capture if they are clusters (Sagey 1986; Rose and Walker 2004; Danis 2017). As is sometimes the case in languages

[7] 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) and Danis (2017) (Maes writes it as ŋb). Note that the Ngbaka we discuss in this subsection is a distinct language from Ngbaka Ma'bo, the latter of which has been described by Thomas (1963) and analyzed by Sagey (1986), Rose and 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).

	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 6: Consonant inventory of Ngbaka, following Maes (1959)

with co-occurrence restrictions, there is an identity exemption: non-adjacent [b. . .b] and [g. . .g] are over-attested in our corpus (Observed/Expected ratios of 3.91 and 4.33, respectively). But there are restrictions on similar but non-identical stops: [mb] does not occur before [b] (0 occurrences), and [ŋmgb] does not occur before [gb] (also 0 occurrences). It would be difficult to account for these distributions while treating [gb] and [mb] as clusters of singleton [g] and [b].

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

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

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 6, 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 [harkaka:] 'to be rough, stiff' (p. 206).

The computational learner's task is harder for Ngbaka than it is for Fijian, for two reasons. First, the segment /ŋmgb/ requires at least two iterations to be unified. Second, the learner must differentiate the complex segments in Table 6 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 the first iteration, it unified the following sequences: [n d], [g b], [ŋ g], [k p], [n z], [m b], [ŋ m], [v w]. This is almost the right result: the learner does not unify [n w] 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 7 presents the calculations for this first iteration; 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 p(C1C2) of 1.0. We left them out of the table to save room.

The second iteration 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

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 7: Ngbaka inseparability values, first iteration (before unification)

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 8. As before, clusters that are unified are above the line; clusters that remain 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 8: Ngbaka inseparability values, second and third iterations

The learner thus succeeds in addressing the challenges posed by the Ngbaka data. 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.

3.1.2 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 9.[8] 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 9: 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 (10f–g) to the repaired cluster in (10h)). The examples in (10) also illustrate other aspects of Mbay phonotactics: nasal-stop sequences 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 (10c–e).

(10) Mbay phonotactics (Keegan 1997 pp. 2–11)

a.	dāŋ	'misery'	e.	kùndə́	'millet drink'
b.	nàr	'money'	f.	sèrbétè	'towel' (Fr. *serviette*)
c.	ndà	'hit'	g.	làmpɔ̋ɔ̏	'taxes' (Fr. *l'impôt*)
d.	ǹda	'he show'	h.	pə̀lə̀r	'flower tree' (Fr. *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 come up with (i.e., sonorants are allowed in coda position, and both sonorants and obstruents are allowed in onset position).

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).

[8]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.

The calculations for Iteration 1 are presented in more detail in Table 10, which demonstrates the large gap between the inseparability measures of nasal-stop sequences and other clusters.

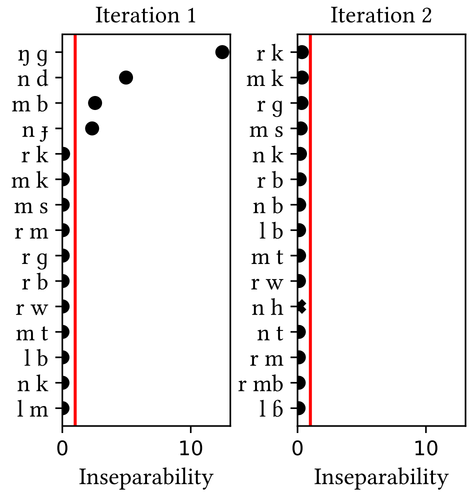

Figure 2: Top 15 inseparability values for various CC sequences at iterations 1 and 2 (Mbay)

	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 10: 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 7), 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.

3.1.3 Turkish and Hebrew affricates

Here, we briefly sketch two additional cases where the arguments for complex segments are fairly clear: 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 and Kerslake 2004; Kornfilt 2013). Phonotactically, Turkish is a CVC(C) language, meaning that CC clusters are allowed word-finally and medially, but not initially. As shown in (11), 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] 'jazz' not [diʒas] (11c)). While the distribution of [dʒ] is restricted compared to [tʃ] (see §2.7), it still patterns more like a segment than a cluster.

(11) Turkish phonotactics (from Göksel and 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, Inkelas et al. 2000; TELL includes paradigms; we used all wordforms but the results are equivalent with only 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.

In Hebrew, the one complex segment is [ts]. The arguments for this analysis are laid out in Bolozky (1980). Hebrew allows a range of clusters in word-initial and medial position, but initial clusters can be at most two consonants (with rare exceptions, like [skleʁozis] (12q), attested only in loanwords). With respect to this restriction, [ts] functions as a single segment: words like [btsalim] 'onions' (12f) and [tsdaka] 'charity' (12g) are licit. 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 and Bat-El 2019, Asherov and Cohen 2019).

(12) Hebrew phonotactics (from Asherov and Bat-El 2019; we follow them in ignoring voicing assimilation)

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 and Becker 2006). On the first iteration, the learner identified [t s] as a segment (insep. 1.74); the runner-up, [d ʒ], was nowhere near the threshold (insep. 0.26). The second iteration found no further complex segments, leaving a total of 297 clusters non-unified.

3.1.4 Quechua: the nature of the learning data

We conclude this section with Bolivian Quechua (Parker and Weber 1996; MacEachern 1997, Gallagher 2011; 2013; 2016a). Quechua presents clear analytic arguments for affricates, which are supported by experimental evidence. Moreover, Quechua supplies an answer to the question from §2.8: complex segment representations are learned from type frequencies in morphemes, not phonological words.

Arguments for a complex segment analysis Quechua offers several arguments for the affricates [tʃ, tʃ', tʃʰ]. First, analyzing these as affricates makes sense given its obstruent inventory (see (13)). There is a three-way laryngeal contrast in stops: plain, aspirated, and ejective. Classifying [tʃ, tʃ', tʃʰ] as affricates reinforces this symmetry. If [tʃ, tʃ', tʃʰ] were analyzed as clusters, however, the inventory would have to contain [ʃ', ʃʰ], which would occur only after [t] and never by themselves.

(13) Quechua obstruents (Gallagher 2019:40)

	labial	dental	post-alveolar	velar	uvular	glottal
stops	p pʰ p'	t tʰ t'	tʃ tʃʰ t'	k kʰ k'	q qʰ q'	
fricatives		s	ʃ	x		h

There are also phonotactic arguments. Assuming affricates, Quechua is a CVC language that imposes simple constraints on medial CCs: fricatives and sonorants can occur in coda, but stops cannot (Gouskova and Gallagher 2020:87). Affricates pattern in every way with stops: they may occur word-initially and medially after certain consonants, but they are not allowed in coda (see (14)). If the affricates were stop-fricative clusters, Quechua phonotactics would be much more convoluted, with statements such as, "CCC clusters are not allowed unless the medial C is [t], and the last C is [ʃ, ʃʰ, ʃ']."

(14) Quechua phonotactics exemplified (Gouskova and Gallagher 2020:87, Gallagher 2016b, Laime Ajacopa 2007)

a.	t'impu	'boil'	d.	tʃ'uspi	'fly'	g.	tʃixʎa	'choose'
b.	ʎaŋk'a	'work'	e.	kʰutʃi	'pig'	h.	q'oɲtʃa	'oven'
c.	wajk'u	'cook'	f.	tʃawpi	'middle'	i.	ɾixtʃ'ari	'to get up'

Quechua also supplies an argument from non-local laryngeal co-occurrence restrictions, which are studied in detail by Gallagher 2010 *et seq.* Quechua restricts aspirated and ejective stops as follows. First, they do not occur in suffixes at all (Parker and Weber 1996; Gouskova and Gallagher 2020). Second, while allowed in non-initial positions in roots, neither ejectives nor aspirates may be preceded by other stops—plain or otherwise (see (15)). There is a number of ways to analyze this, but grouping stops and ejectives into a natural class is crucial. Consider what it would take to capture this distribution if [tʃʰ, tʃ'] were analyzed as clusters instead. One would have to say that [ʃʰ, ʃ'] can (in fact, must be) immediately preceded by stops, since they always occur right after [t], but they cannot be preceded by stops farther away (*[t̲at̲ʃʰa]).

(15) Quechua long-distance co-occurrence restrictions (Gallagher 2019:40, 2016b: 106)

a. Ejectives and aspirates in initial position

tʃʰukuj	'swaddle'	pʰawaj	'to run'	✓tʃʰ. . .k	
k'atʃa	'pretty'	qʰaɾi	'man'	✓k'. . .tʃ	

b. Ejectives and aspirates in medial position; no preceding stops

ɾit'i	'snow'	mosqʰoj	'to dream'	*t. . .qʰ	*t'. . .qʰ
satʃ'a	'tree'	ʎimpʰi	'color'	*q. . .tʃ'	*tʃʰ. . .k'

Since this aspect of Quechua is well-studied, we have experimental evidence that affricates and stops pattern together from behavioral data (Gallagher 2019). Gallagher investigated the co-occurrence restrictions using a nonce word repetition task (see §6), which demonstrated that Quechua speakers repeat phonotactically legal wugs (e.g., [t̲'aʎwa, k̲'ap̲u]) 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—i.e., is *[k̲at̲ʃ'a] as bad as *[t̲ant̲'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.[9] This suggests that the non-local restrictions described in (15) hold irrespective of the simplex vs. complex status of the relevant segments.

Together, these three considerations (inventory structure, local cluster phonotactics, non-local consonant phonotactics and the psychological reality thereof) strongly suggest that Quechua speakers treat the postalveolar stop-fricative sequences as single segments.

Learning simulations and the nature of the learning data For our simulations, we used three kinds of training data:

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

(ii) a tokenized list of 1,484 suffixes and roots from a morphologically segmented newspaper corpus (described in Gouskova and Gallagher 2020),

(iii) a list of 10,847 phonological words from the newspaper corpus (Gouskova and Gallagher 2020).

When trained on (i) and (ii), the learner found the target inventory [tʃ, tʃ', 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 [tʃ'] (180) and aspirated [tʃʰ] (74). The plain affricate is a lot more frequent.

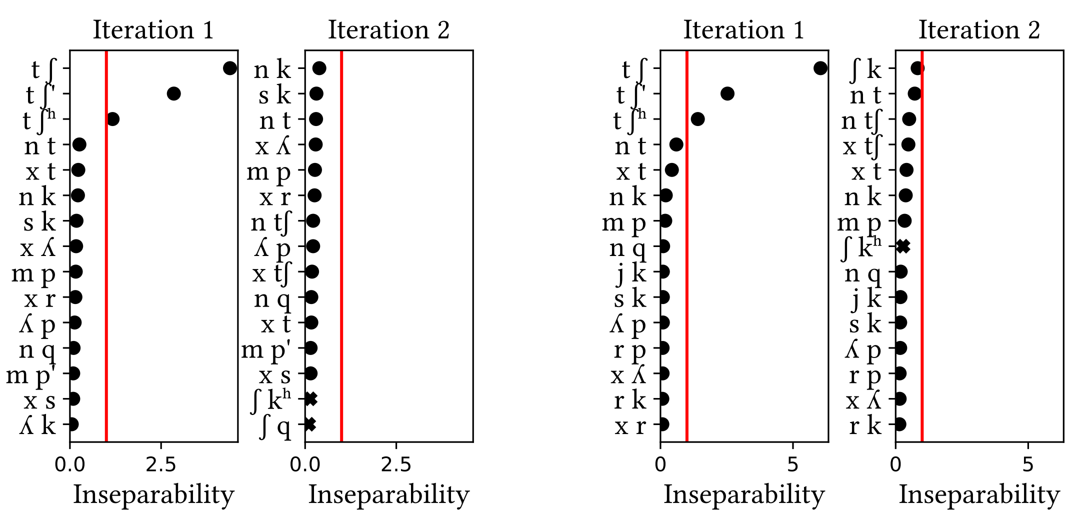

Figure 3: Quechua simulations, roots (left) and morphemes (right)

By contrast, the learner does not do well when trained on Quechua words, (iii). The learner runs as

[9]The ANOVA model comparison assessed a model 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 (χ^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).

many as nine iterations, unifying [tʃ] and [sq], then [ɲtʃ], [sk], [jk], and [tʃ'], [rq], and so on. It does end up finding the three affricates, but it also unifies all sorts of other sequences. The reasons for this failure become clear when we look at where the most "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* '1pl. excl.', *-rqa* 'past'). Quechua is exclusively suffixing, so when its roots combine with such suffixes, the result is CVC syllables, e.g., [ʎaŋk'a-rqa-ŋki] 'work-past-2sg', [puri-spa] 'walk-gerund', [hamu-sqa-jki-ta] 'come-part-2sg-accusative'. The second reason is the distributional restrictions on stops: (i) ejectives and aspirates (including [tʃʰ, tʃ']) do not occur in suffixes, (ii) neither plosives nor affricates occur in codas, (iii) aspirates and ejectives do not occur when preceded at any distance by any other stops. All of this results in [tʃ] being common (3,494 occurrences) and inseparable (5.32), but its ejective and aspirated counterparts are less common and less inseparable than certain clusters that occur in common suffixes. Training the learner on morphologically complex words in such a language makes it inevitable that it will unify the wrong things.

This suggests that attending to frequencies in a list of words is the wrong strategy for a language like Quechua; the distribution of complex segments must be learned from a more abstract dataset. This was not necessary 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 to the affricate inventory than the learner of Turkish? We return to this issue when we consider Russian in §3.2.2.

3.2 Adjudicating between complex segments and clusters

The second set of case studies includes languages where complex segments have been posited but are more controversial. We discuss three such cases here: Latin [k w] and [g w] (§3.2.1), Russian [t s] and [t ɕ] (§3.2.2), and English [t ʃ] and [d ʒ] (§3.2.3). Discussion of a fourth case that falls into this category, Modern Greek [t s] and [d z], is postponed until §5.3.2. Unsurprisingly, given the unclear phonological status of these sequences, the learner finds complex segments in some cases and clusters in others.

3.2.1 Latin [k w] and [g w]

The consonant inventory of Classical Latin in Table 11 is adapted from McCullagh (2011:84).[10] Our interest is in the sequences [k w] and [g w], 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 11: Classical Latin consonant inventory

There are arguments for a complex segment analysis of [k w] and [g w], but they are not convincing, as discussed in detail by Devine and Stephens (1977: Ch. 9). One argument is that [k w] and [g w] are the only stop-[w] clusters in Latin (no [p w, t w, d w], etc.). As Devine and Stephens point out, this does

[10]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.

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 data from a number of languages, including 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 §4.4). A second argument is that the Roman grammarians treated [k w] and [g w] as segments, so we should too. As Devine and Stephens (1977: Ch. 4) carefully lay out, however, the segmental status of [k w] and [g w] has likely been debated since late Republican times. A third argument often given in favor of monosegmental [kw] is that it consistently did not make position in Latin poetry (Devine and Stephens 1977:51-68; see also McCullagh 2011 for a summary). 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 [k w] is a segment, since there are many other reasons why [k w] might metrify differently from other clusters.[11] In sum, every argument for treating [k w] and [g w] 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 (https://linguistics.ucla.edu/people/hayes/learning/latin.zip), Whitaker’s online dictionary (http://mk270.github.io/whitakers-words/, 84,000 words), and Lewis et al. (1969), with 49,725 words. The results were qualitatively the same: no complex segments; [n t] comes in as the most inseparable, but never passes the threshold. 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.

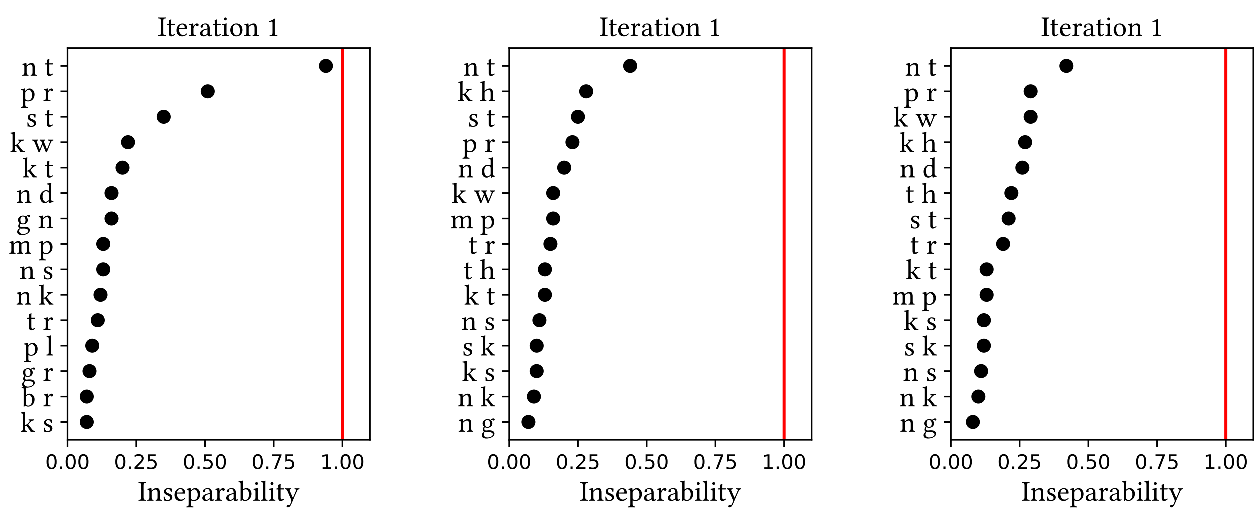

Figure 4: Latin simulations: nouns (left), Whitaker (center), Lewis et al. (right)

Our results suggest that [k w] and [g w] were clusters in Classical Latin. Latin is also interesting for a broader reason: in the case studies up to this point, the learner treated many reasonably frequent consonant

[11] 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).

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, discussed briefly in §5.3.2, and French, in §6; the latter uncontroversially lacks complex segments).

As anticipated in §2.7, 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 2007:3.2.1). (This is an oversimplification, since palatalization applied in other contexts and not necessarily at the same time, but it suffices for demonstration). We simulated these sound changes by replacing the affected sequences in our Latin data. When we trained the learner on pseudo-Italian, it unified all four sequences in two iterations, treating them as affricates (in paradigms and Whitaker; in the Lewis/Short simulation, [t s] did not pass the threshold on 2nd iteration but was on top). Fig. 5 shows that [dz, tʃ, dʒ] pass the threshold on the first iteration, and [ts] on the second, once [tʃ] has been unified.

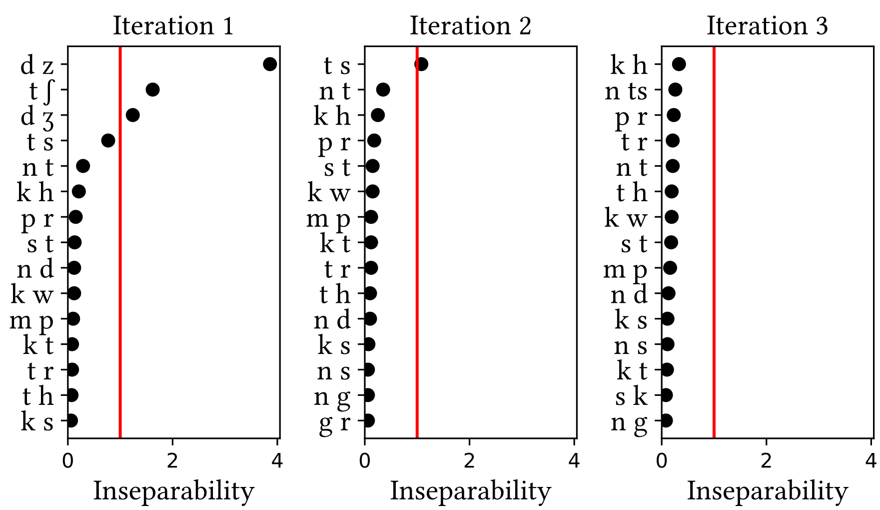

Figure 5: Simulation on Latin (Whitaker dataset) with pseudo-Italian changes applied {k , g, t, d} →{t s, d z, t ʃ, d ʒ} /___i

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.

3.2.2 Russian affricates

The traditional analysis of Russian posits two affricates: [ts] and [tɕ]. As we show below, 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 basic result, we revisit the issue of the nature of learning data: when the learner is trained on connected speech or morphemes, it unifies [tɕ] but not [ts].

The inventory we assume (following Padgett 2003; Padgett and Żygis 2007) is given in Table 12.[12]

[12]Note that we did not give the learner a chance to consider Russian palatalized Cs for unification. This is because it is not clear to us how to transcribe the distinction between C^j and Cj in the initial state: Russian has contrasts such as [lʲot] 'ice' ∼[ljot] 'pours', [abjom] 'volume' ∼ [grʲibʲom] 'we row', so transcribing all as [C j] would neutralize this distinction. (One possibility

	labial	dental	(alv-)palatal	retroflex	velar
stops	p, b, pʲ, bʲ	t, d, tʲ, dʲ			k, g, kʲ, gʲ
affricates		ts	tɕ		
fricatives	f, fʲ, v, vʲ	s, z, sʲ, zʲ	ɕː	ʂ, ʐ	x, xʲ
nasals	m, mʲ	n, nʲ			
liquids		l, lʲ, r, rʲ			
glides			j		

Table 12: Inventory of Russian contrastive consonants

The phonological analysis of [ts] and [tɕ] as affricates in Russian is neither questioned nor supported by argumentation in most sources. This could be for two reasons. First, it is widely accepted in Slavic philology that affricates come from historical single stops (§2.7). Knowledge of this single-segment origin might have biased researchers towards not questioning the affricates' contemporary status. Second, Russian orthography might have obscured the affricates from the linguists' attention: [ts] and [tɕ] are written with single letters, <ц> and <ч> (see §6).

Trubetzkoy (1939) is the exception in considering the status of affricates from several angles. He supplies a phonetic argument, claiming that [ts] and [tɕ] are durationally more similar to simplex segments than to clusters (1939:58). But his intuitions have not (to our knowledge) been supported by any systematic experimental research, and Trubetzkoy casts doubt on his own argument by noting that the durations of simplex segments vary (see §5.3). Trubetzkoy also suggests that [ts] and [tɕ] have the distribution of single segments, since they can occur in word-initial position. But so can many other consonant-fricative sequences in Russian, as shown in (16). Moreover, it is not obvious why the voiceless [ts] and [tɕ] are treated as segments but the voiced stop-fricative sequences are not. Word-initial [dz] occurs in Polish and Belarusian loans (such as (16o)), and [dʐ] occurs whenever [dʒ] is borrowed (16n). Furthermore, the inclusion of just [ts] and [tɕ] leads to a less symmetrical inventory, since singleton stops and most singleton fricatives contrast with their voiced counterparts but affricates do not. Based on Trubetzkoy's symmetrical inventory heuristic, Russian [ts] and [tɕ] might be better analyzed as clusters.

(16) Russian phonotactics

a.	tsɨna	'price'	h.	ksvʲinʲje	'to a pig'	n.	dʐɨnsɨ	'jeans (Eng.)'
b.	tɕuʂ	'nonsense'	i.	kxarʲku	'to a hamster'	o.	dzerʐɨnskʲij	'Dzerzhinsky'
c.	vʲetɕir	'evening'	j.	kxvalʲe	'towards praise'	p.	gʐelʲ	'Gzhel village'
d.	rʲetɕ	'speech'	k.	kfrantsii	'towards France'	q.	vʲitʂatʲ	'to get old'
e.	agurʲets	'cucumber'	l.	pʂɨno	'millet'	r.	imʲitʂ	'image (Eng.)'
f.	tsvʲet	'color'	m.	mɕːenʲijə	'revenge'			
g.	tɕlʲen	'member'						

We can supply (and refute) one more argument for affricates: they alternate with segments, as in [krʲuk] 'hook (sg)' ∼ [krʲutɕ-ja] 'pl', [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 (perf.)' ∼ [pabeʐd-al] 'he won (imperf)'. 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.

would be to transcribe the [j]s with different lengths, so [b j] vs. [b j:], since articulatory studies such as Kochetov 2006:575 find the difference to be one of timing. This would likely result in unification of all palatalized consonants as short [j] would be unattested elsewhere.)

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).[13] The learner unified [tɕ] in the first iteration and [ts] in the second. The results are shown graphically in Figure 6. The more inseparable [t ɕ] (insep= 2.02) occurs 14,248 times; [t] occurs 64,867 times and [ɕ] occurs 18,028 times. On the second iteration, [t s] rises from 0.99 to 1.42; it occurs 17,707 times, with [t] appearing 50,622 times and [s] 56,846 times. Note that even though both subparts of [t s] are frequent, the frequency of [t s] itself is high enough to drive its inseparability up. The next most inseparable sequence was [s kʲ] (inseparability=0.59 on the first iteration, and 0.66 on the second iteration; [s t] rises to 0.93 after [ts] is unified).

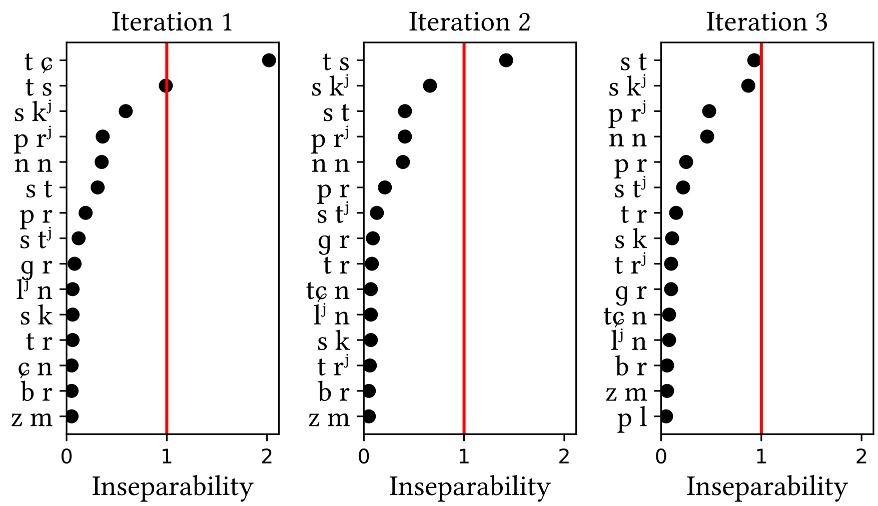

Figure 6: Inseparability for Russian CC sequences

This result suggests that the best argument for the single-consonant analysis of Russian [ts, tɕ] is their distribution, at least in phonological words. While the result is replicated on two training data sets, there are also reasons to be skeptical of it. First, the type frequency of [ts] might be artificially inflated in our data: dictionaries include citation (infinitival) forms of verbs, which often end in [t s a], 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 the connected speech run, inseparability for [tɕ] was 1.89; in the morpheme run, 2.88. There were qualitative differences between these simulations in the runners-up. In the morpheme run, [t s] (0.47) followed [s t] as the second runner-up (0.51); in the connected speech, [t s] was 7th in the last iteration (insep=0.2), trailing clusters in common

[13] We converted the Russian orthography into IPA by script. In addition to <ч>→[t ɕ], the other source of [ɕ] was <щ>. It is usually analyzed as long, as in our Table 12). Transcribing it as [ɕː] would have made it even easier for the learner to find [tɕ], because then short [ɕ] would not occur outside the affricate.

morphemes such as [p r^j].

The divergence between the dictionary simulations and morpheme and connected speech ones is difficult to interpret. There is no clear phonological or behavioral evidence that [ts] is an affricate in Russian. This is equally true for [tɕ], but the latter is much more robustly discoverable in a variety of data. We could certainly question the quality of the mock connected speech dataset and the morphemic lexicon, and trust the phonological word simulations, which find both [ts] and [tɕ]. 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 makes 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.

3.2.3 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 and Halle 1968; Kahn 1976; Selkirk 1982; Borowsky 1986; Moreton 2002; Daland et al. 2011, a.o.). 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 and Halle 1968:223; cf. Jones 1918). The usual phonotactic argument for this asymmetry is that [t s, d z] do not occur word-initially (aside from careful pronunciations of loanwords such as *tsunami*; see, e.g., Ladefoged 1996). Under the traditional analysis, the explanation for [t s] and [d z] not occurring in initial position is that they are clusters, and stop-fricative clusters are not allowed in initial position. But this is unsatisfying, as any characterization of English phonotactics must include statements about individual singleton consonants being banned from initial position: it could be the case that [ts] and [dz] are single segments but banned from initial position, just like [ŋ]. The traditional analysis also has difficulty explaining why [tʃ] cannot combine with other consonants in initial position. 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.

We tried two corpora: Celex (Baayen et al. 1993, 72,969 words) and the Carnegie Mellon Dictionary (version of Hayes and 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 *cour<u>tsh</u>ip,* but postalveolar-postalveolar [c ʃ] in *ke<u>tch</u>up*. 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. 7).

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 13). Note also that the cross-morpheme and non-homorganic [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.

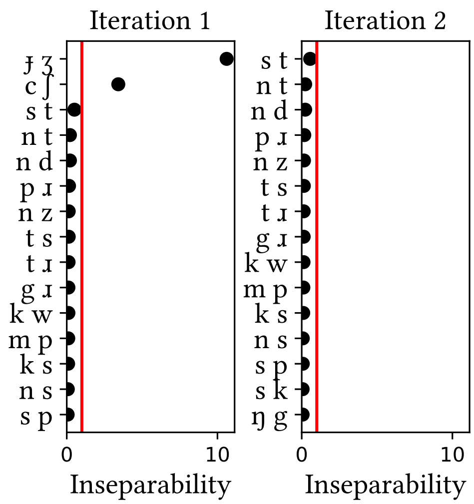

Figure 7: English inseparability measures, affricates transcribed narrowly

	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 13: 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 14). 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 14: 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. But, our learner is clear on the status of [t s] in English: it is a cluster, not an affricate.

3.3 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 (1957) provides a CV representation of [sunda] as CVCCV and [ŋimpi] as CVCCV (p. 89), and refers to them as sequences (his fn. 1). Cohn (1992) does not include them in her posited inventory, and describes nasal-stop sequences as split across a syllable boundary (p. 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 15. 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 included 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 15: 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 and Sastra Sunda (1985), a monolingual Sundanese dictionary (16,327 headwords entered by hand). In addition to the segments in Table 15, 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. Seven sequences qualify for unification: [ɲ c], [n d], [ɲ ɟ], [m b], [m p], [n t], and [ŋ k]. The learner unifies these sequences, and runs the procedure again. On the second iteration, the eighth and only remaining nasal-stop sequence, [ŋ g], now qualifies for unification. All other clusters fall below the threshold. 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 8.

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 2016:5), but phonotactic restrictions on initial segments are not unheard of (e.g., English [ŋ]). The learner's conclusion thus mirrors descriptions that treat the voiced series as

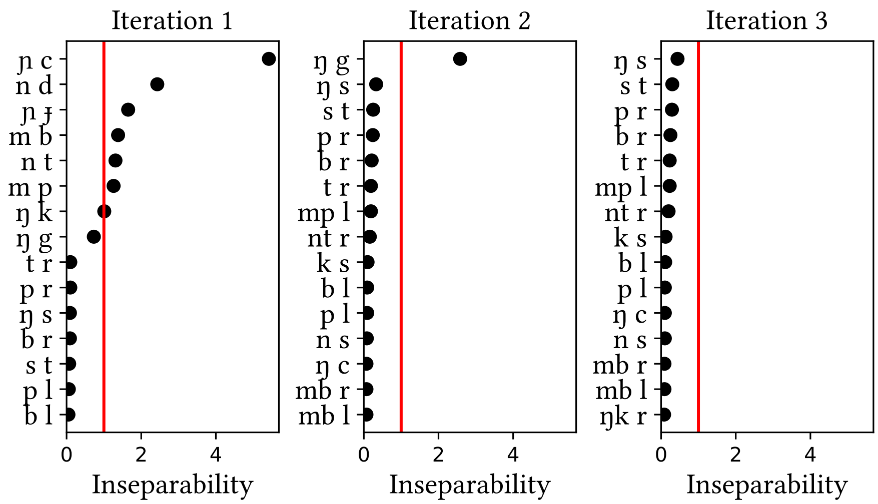

Figure 8: Sundanese simulation

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 by Stanton (2016:1091), who notes that Makaa (among other languages) is recorded as having voiceless prenasalized /mp/ but not /p/ as part of its inventory (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 (pp. 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 and Stivers 1996, Pater 1999, and our S4.4). 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 (2016) observation that "the distribution of NDs completely parallels that of NTs" (p. 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.

4 Typology

Coupled with additional assumptions, our proposal makes predictions for the typology of complex segments. We mainly focus on generalizations about their size, as these have been addressed by other proposals (§4.1), and the typology of complex segment size and cluster size is well understood (§4.2). Section 4.3 is a case study of Shona, which is typologically unusual in allowing four-part complex segments. Section 4.4 briefly discusses several generalizations about the composition of complex segments.

4.1 Theories of limitations on complex segment size

Typologically, complex segments are often composed of two subparts (e.g. [mb], [ts]), less commonly three (e.g. [ɲdʒ]), and rarely four (e.g. [ɲdʒw]). Some proposals capture these generalizations 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 (e.g. Inkelas and Shih 2013, 2016; Garvin et al. 2018; Shih and Inkelas 2018, 2019, cf. Schwarz et al. 2019) imposes similar limitations on the size of complex segments. In Q theory, each segment consists of sequenced subsegments. Most work on Q theory assumes that there can be at most three subsegments: "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"

These proposals capture limitations on the size of complex segments by stipulation: there is 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 on the basis of 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.

4.2 Rarity of long consonant clusters

Typologically, the bigger the consonant cluster, the less common it is. Gordon's (2016:91) 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 two-membered onsets and two-membered codas, for example, the maximum cluster size will be four (V<u>CC.CC</u>V). The number of languages per predicted maximum cluster size, given Gordon's survey data, is in (17). A minority (30/97) are predicted to allow 4-membered or longer clusters.

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

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

Large consonant clusters are rare not only cross-linguistically but also within languages. Even for the 30 languages in (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 (see 3.2.2) contains 289,830 intervocalic consonant sequences (assuming that affricates are complex segments, not clusters). Russian has clusters up to five Cs, but they only occur 31 times in our corpus. As is evident from (18), single consonants and CC clusters are far more common.

(18) Frequency of intervocalic consonant sequences in Russian

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

A 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.[14] The rest of the languages in Rousset's study make the same point; see her p. 116 for details.

4.3 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 16.

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

Table 16: Zezuru: simplex consonants (Fortune 1980)

According to Fortune (1980) and others, the basic phones can combine into affricates [pf, bv, ts, dz, tʃ, dʒ, ts̫, dz̫], prenasalized consonants [mb, nd, nz, ŋg, . . .], and velarized consonants [tw, dw, 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; two examples are [ma-girazi] from English 'glasses', [mu-puraŋga] from Portuguese *prancha* 'gum-tree').

To see if our learner would find these longer complex segments, we trained it on Duramazwi ReChisona, an electronic dictionary (15,830 entries, Chimhundu 1996). We converted the Shona orthography into Zezuru transcriptions following Fortune (1980). Over four iterations, our learner finds many complex segments (41 total). We only show the counts for four- and five-consonant sequences in (19). The learner unifies all but [ɲ d ʒ g w]. Its inseparability is above 1 on the final iteration, but its frequency is indistinguishable from 0.

(19) Counts for Zezuru four- and five-part segments

Sequence	Count	Sequence	Count
[ɲ d ʒ g]	42	[t ʃ k w]	12
[t s k w]	37	[d ʒ g w]	6
[d z g w]	26	[ɲ d ʒ g w]	1
[n z g w]	25		

[14] 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.

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, as is the common assumption). 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.

4.4 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] (§3.2.1), and (2) there is a dispreference for voiceless nasal-stop sequences (§3.3). 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 and Lorentz 1977). When languages have labialized consonants, they will often have a gap of precisely the same combinations that are ruled out as clusters in other languages. For example, 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 no 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 for nasal-voiceless-stop (NT) clusters (Pater 1999; Hayes and Stivers 1996, et seq.). For segments, this dispreference manifests in the rarity of voiceless prenasalized stops. Maddieson and 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[15]—we would expect NT to be unified less frequently.

Other links between the typologies of prenasalized stops and nasal-stop clusters might also be explained this way. For example, the vast majority of prenasalized consonants are homorganic, which can be linked to the common requirement that a nasals assimilate in place to a following consonant (Mohanan 1982; Ito 1986; Padgett 1995b, and many others). This requirement holds as a statistical trend even in languages that allow both heterorganic and homorganic nasal-stop sequences—in languages like this for which we have corpora (Yindjibarndi, Wargamay, Russian), the homorganic sequences are more frequent. In Yindjibarndi, Wordick's lexicon contains a total of 574 homorganic nasal-stop clusters and 160 heterorganic clusters (Stanton 2019). Dixon (1981:23) reports that in Wargamay, 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 our corpus) and [m p, m p^j] (548 occurrences) than heterorganic ones such as [m k, m k^j] (197 occurrences).

On a more general note, constraints on consonant sequencing could explain a number of generalizations about the typology of complex segments. For example, [nd] and [kw] are fairly frequent in the

[15] 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 in English, other things are at play: for example, [m p] is allowed word-finally, but [m b] is not (Kaplan 2007).

inventories of the worlds' languages, but [nl] is—to our knowledge—unattested (Maddieson and 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. In this way, 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?

5 Alternatives

We have advocated for an approach in which learners posit complex segments on the basis of the transitional probabilities of consonant sequences. This section discusses three alternatives. First (§5.1), we defend our decision to calculate probabilities over segments rather than natural classes. Then (§5.2), we discuss the possibility that learners construct complex segments so as to simplify the phonotactic grammar, and show that such an approach makes incorrect predictions regarding the segmental inventories of English and Russian. Finally (§5.3), we explain why we doubt that learners use phonetic information as a general strategy for identifying complex segments.

5.1 Inseparability over natural classes

We treat discovering complex segment representations as a problem for the learner that already has segments. Inseparability is calculated over segments, in order to create shortcut representations for certain sequences that appear to have the distributions of units. Moreover, while newly unified complex segments are given featural representations, these representations do not directly inform or limit the learner while it is calculating inseparability. There is no requirement, in other words, that segments being unified must match in place of articulation, voicing, or any other features; wherever such restrictions hold, they come from general constraints on sequencing. Here, we consider an alternative where inseparability is calculated over natural classes.[16] The upshot is that this version of the learner misses generalizations where present, and the change in the combinatorics makes thresholds nonsensical and the results of the learner difficult to interpret, both within and across languages.

Intuitively, there is much to recommend an approach based on natural classes. First, it is in line with other research on statistical phonological learning, which assumes that generalizations are made over features and natural classes in addition to segments (Albright and Hayes 2003; Hayes and Wilson 2008; Albright 2009; Adriaans and Kager 2010; Gouskova and Gallagher 2020). To the extent that these models have access to segmental representations, segments are simply natural classes of one (e.g., these models represent [s] as the collection of feature values that uniquely picks out that segment [-son, +cont, +strid, COR, +anterior]). Second, and perhaps more importantly, a calculation over natural classes might allow the learner to discover segments such as [tʃ] and [dʒ] in English on a single pass, just in case the natural class bigram [-son,-cont] [+strid, -anterior] is highly inseparable.

To explore this alternative, we created natural class representations from feature charts defining all the segments, along the lines of Hayes and Wilson's (2008) learner. As we did for our segment-based learner, we restricted the class-based learner's calculations to sequences of natural classes that contained only consonants. We also did not iterate the learning procedure, for reasons that will become clear shortly.

5.1.1 No clear threshold

The first point of difference between our proposal and the natural class alternative is that the latter does not allow us to establish a clear threshold for deciding what sequences should be unified, either within

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

or across languages. This has to do with the combinatorics of natural classes (Hayes and Wilson 2008, Gouskova and Gallagher 2020). The number of natural classes in any language considerably exceeds the number of segments, so the search space for a class-based learner is much bigger than the search space for a segment-based learner. Comparisons for a representative set of languages are given below. Table 17 shows totals for simplex segments our learner sees in the first iteration. The natural classes column includes all the classes definable using the feature set we used (including vowels and consonants). The column "CC bigrams" shows counts for sequences of [-syll] *segments* attested in the initial dataset on the first iteration. The third column shows these same attested bigrams, but counted in terms of the *natural classes* they belong to; for example, [m b] is counted multiple times as [+nas] [-son], [+LAB] [+voice, +LAB, -cont], etc. The rightmost column shows the maximal inseparability values calculated over natural classes as opposed to segmental bigrams.

Language	Segs	Nat. Cl.	CC 2grams	Nat. Cl. 2grams	Max N.Cl. insep.
English (Celex, broad)	36	199	340	7,825	0.05
English (Celex, narrow)	38	214	371	10,430	0.0258
Fijian	25	136	6	2,046	0.081
Greek	30	131	181	8,052	0.00365
Hebrew	31	190	272	13,212	0.0032
Latin	32	159	178	12,463	0.0017
Ngbaka	33	183	28	4,350	0.0212
Russian (Zaliznjak)	41	256	597	45,059	0.0017
Quechua (roots)	33	141	124	5,867	0.0215
Turkish	38	194	315	8,074	0.0151

Table 17: Segment vs. natural classes, by language

The increase in computational demands is not nearly as problematic as the increase in the number of sequences over which probability has to be calculated. Calculating over thousands, as opposed to dozens or hundreds of bigrams, reduces the probability of any one bigram so much that none even approximate an inseparability of 1 in any of the languages. This makes the notion of a threshold no longer tenable. Furthermore, the values calculated over natural classes are both small and not in an obvious way comparable between languages. By contrast, our learner makes its decisions based on simple reasoning that applies to all languages: any sequence occurs more often together than apart (i.e., the ratio exceeds 1) is unified.

5.1.2 No obvious negative results

The lack of a clear threshold makes it difficult to know when the learner yields a negative result. For example, in our Latin case study (§3.2.1), we concluded that the phonotactic arguments for [kw] and [gw] were weak, and our a segment-based learner did not identify any unifiable sequences (the most inseparable sequence for Latin was [n t]). When we calculated inseparability for Latin natural classes, however, the most inseparable sequence was [+cons, -long] [-long], which covered sequences of non-geminate consonants ([b|d|f|g|h|k|l|m|n|p|r|s|t b|d|f|g|h|j|k|l|m|n|p|r|s|t|w]). The inseparability value for these was 0.0017. This is the same value as in Russian, where the top most inseparable sequence was correctly identified as [tɕ]. Compare this to Greek (§5.3.2): the phonotactic arguments are inconclusive, just as in Latin, but the most inseparable sequence is [+cons,-nas,-syll] [+cons], with a value of 0.0036. The natural class bigram expands to [b|c|d|f|g|j|k|l|p|r|s|t|v|x|z|ç|ð|ɣ|ʎ|ɟ|θ b|c|d|f|g|j|k|l|m|n|p|r|s|t|v|x|z|ç|ð|ŋ|ɣ|ɲ|ʎ|ɟ|θ]. The Greek natural class sequence has a much higher inseparability value than that of Russian [tɕ] or Latin non-geminate clusters. The natural class calculation for Hebrew identified [t s] as most inseparable, with a value of 0.0032 (lower than Greek). So how should Greek, Latin, Hebrew, and Russian be treated? Should the learner unify the topmost natural class sequence and iterate, or should it posit a cut-off based on some language-internal criterion? The answers to these questions are not clear to us.

5.1.3 Missing actual natural class-based generalizations

Even in cases where one might expect the natural class learner to find a clear generalization, it fails. Consider Fijian. In our segment-based calculation (§2.3), the learner found all the prenasalized stops and [tʃ] on the first iteration. It required a second iteration only for the three-member cluster [ndʒ] and the prenasalized trill [nr]. When counted as natural classes, prenasalized stops have an inseparability value sandwiched between [ŋ g], [m b] on the one hand and [n d] on the other (shown in Table 18). The affricate natural class bigram, [d|t ʃ|ʒ], is also nowhere near the top in this calculation. This is an odd outcome that arises because each segment is a natural class—and yet surely any phonological theory should be able to refer to individual segments as well as larger natural classes (Albright 2009, a.o.).

featurally defined nat. class bigram	as segments	Insep	N(Cl1 Cl2)
[+dor,+nas] [+dor,+voice,-son]	[ŋ] [g]	0.081	1026
[+lab, +nas] [+lab, +voice, -cont, -son]	[m] [b]	0.0661	2328
[+nas] [+voice, -cont, -son]	[m|n|ŋ] [b|d|g]	0.0649	5866
[+ant, +nas] [+ant, +cor, +voice, -cont, -son]	[n] [d]	0.0591	2512
[+dor,+voice,-cont] [+dor,+voice,-son]	[ŋ|g] [g]	0.0559	1026
[+dor,+son] [+dor,+voice,-son]	[w|ŋ] [g]	0.053	1026
[+lab,+voice,-cont] [+lab,+voice,-cont,-son]	[b|m] [b]	0.0484	2328
[+voice,-cont] [+voice,-cont,-son]	[b|d|m|n|ŋ|g] [b|d|g]	0.0477	5866
[+ant,+cor,+voice,-cont] [+ant,+cor,+voice,-cont,-son]	[d|n] [d]	0.0445	2512
[+lab,+nas] [+lab,-cont,-son]	[m] [b|p]	0.0444	2328
[+lab,-cont] [+lab,+voice,-cont,-son]	[b|m|p] [b]	0.0427	2328
[+ant,-cont,-voice] [-ant,-voice]	[t] [ʃ]	0.0427	1985
[+cons,+son] [+voice,-cont,-son]	[l|m|n|r|ŋ] [b|d|g]	0.0387	5866
[+ant,-cont,-son] [+strid,-ant]	[d|t] [ʃ|ʒ]	0.0381	2305

Table 18: Top most inseparable sequences in Fijian, as natural classes

We also identified at least one pathological case where making generalizations at the level of natural classes is incorrect, because the set of complex segments cannot be described in a way that makes the right cut. For Russian, our segment-based learner identified [tɕ] and then [ts], in two iterations (§3.2.2). The natural class-based learner also identifies [tɕ] as having a high inseparability value (0.0017), but it also identifies [t] [ɕ|s|sʲ|ʂ] as having this same inseparability value. The inclusion of [sʲ] and [ʂ] is problematic, since [t sʲ] is only possible at morpheme boundaries (in words like /ot-sʲel/ [otsʲel] 'sat away from'), and would be analyzed as a cluster in any mainstream treatment of the language. Similarly, [t ʂ] is almost exclusively a heteromorphemic cluster; hardly a candidate for complex segment status. The learner that looks at natural classes rather than segments would incorrectly unify these sequences. This happens because affricates do not really constitute a clear natural class to the exclusion of other stop-fricative sequences in Russian: not all stops followed by a coronal stridents constitute complex segments. The gapped, asymmetric nature of the Russian affricate inventory requires a segment-based level of analysis.

featurally defined nat. class bigram	as segments	Insep	N(Cl1 Cl2)
[t ɕ]	[+back,-strid,-voice] [-ant,-back,-voice]	0.0017	12900
[t ɕ|ʑ]	[+back,-strid,-voice] [+strid,-ant,-back]	0.0017	12900
[t s|sʲ|ɕ|ʂ]	[+back,-strid,-voice] [+strid,-voice]	0.0017	28049
[t f|fʲ|s|sʲ|x|xʲ|ɕ|ʂ]	[+back,-strid,-voice] [+cont,-voice]	0.0015	28171
[d|t ɕ|ʑ]	[+back,-strid] [+strid,-ant,-back]	0.0014	12905
[d|t ɕ]	[+back,-strid] [-ant,-back,-voice]	0.0014	12900
[d|t s|sʲ|ɕ|ʂ]	[+back,-strid] [+strid,-voice]	0.0014	28049
[t sʲ|ɕ]	[+back,-strid,-voice] [+strid,-back,-voice]	0.0014	12984
[t s|sʲ|z|zʲ|ɕ|ʂ|ʐ|ʑ]	[+back,-strid,-voice] [+strid]	0.0013	28049
[t ɕ|ʂ]	[+back,-strid,-voice] [-ant,-voice]	0.0013	13070
[t sʲ|zʲ|ɕ|ʑ]	[+back,-strid,-voice] [+strid,-back]	0.0012	12984
[t fʲ|sʲ|xʲ|ɕ]	[+back,-strid,-voice] [+cont,-back,-voice]	0.0012	13001
[d|t f|fʲ|s|sʲ|x|xʲ|ɕ|ʂ]	[+back,-strid] [+cont,-voice]	0.0012	28174

Table 19: Top most inseparable sequences in Russian, as natural classes

We should emphasize that in many cases, the natural class bigram calculations converge on similar results that we achieve with segments—primarily because each segment can be defined in terms of features, and the learner identifies bigrams of natural classes with just one member each as most inseparable quite often. For example, [d ʒ] is the most inseparable sequence in both Turkish and English (broadly or narrowly transcribed). The problem is that the natural class-based version rarely finds the right generalization in cases where it would be desirable. It does so in Quechua (when trained on roots): the topmost sequence is [t ʃ|ʃʰ|ʃ'], precisely the right result. We doubt that the success in Quechua is anything to celebrate, as it is something of an anomaly: the learner otherwise fails to find class-based generalizations that covers the right set of segments.

5.2 The phonotactic alternative

Phonological arguments for complex segments often rest on phonotactic argumentation—we have detailed such arguments throughout our case studies. Here, we consider the possibility that learners use phonotactics to decide which sequences are complex segments (as hypothesized explicitly by Herbert 1986). Perhaps the learner tries a phonotactic grammar with the cluster representation and a phonotactic grammar with the complex segment representation for each candidate sequence, and evaluates the fit to see if there is an improvement. This strategy does not work, for a fairly intuitive reason: phonotactic grammars are always improved when they have access to shortcut representations for certain sequences. We focus in this section on what the phonotactic procedure might look like for English and Russian. For these two languages, the phonotactic strategy leads to the questionable conclusion that they have all sorts of complex segments that have never been posited for them.

Devising a phonotactics-based strategy for learning complex segments requires nontrivial decisions about structuring the hypothesis space. Since the learner does not know in advance what types of complex segments its language could have, it would need to navigate through a lot of possibilities. Does English have prenasalized stops in words such as *bingo* and *gumbo*? Should the learner consider labiovelars, such as [kp, gb] in *jackpot* and *rugby*? Since English has potential affricates ([tʃ, dʒ, ts, dz]), does the learner attempt a complex representation for all of them, or does it try one at a time? If they are tried in order, how is the order determined? What about possible three- and four-part complex segments (e.g., [ntʃw, ndʒw] in *inch worm* and *binge-watch*)?

We set those questions aside, and tested the phonotactics-based strategy by manually creating progressively more elaborate representations, and training the UCLA Phonotactic Learner (Hayes and Wilson 2008; Wilson and Gallagher 2018; Gouskova and Gallagher 2020) on the resulting datasets. We used the gain-based version of the learner rather than the 2008 Observed/Expected version; see Gouskova and Gallagher (2020) for details. In order to assess the resulting phonotactic grammars, we needed a measure of best fit. We used two such measures: (i) the **log probability** of the data given the grammar induced by the learner; and (ii) a **generality measure**, which counts up how many segments, on average, each constraint in the grammar refers to. Log probability is calculated by the learner for the entire grammar after each constraint is added (see Hayes and Wilson 2008:386–387); we use the final, highest value. Generality seems to us to correlate with a feature of good phonological grammars: their constraints are maximally general. For example, in Fijian, being able to refer to complex consonants allows the phonotactic grammar to make a simple generalization: [-syll][-syll] sequences are not allowed. If these segments were not represented as segments, the learner would have to induce constraints against all the un(der)attested consonant combinations, and therefore its constraints on average would be less general.

We tested this approach on English, Russian, Fijian, Ngbaka, and Mbay. In all the languages, adding the linguist-posited complex segments improves the fit of phonotactic grammars, compared to representing the sequences as clusters. Thus, in English, representing [tʃ, dʒ] as affricates allows the learner to posit more general constraints and to describe attested vs. unattested sequences with a higher log probability. For Fijian, the improvement is striking and pronounced; the more complex the segments in the learning data, the better the fit and the simpler the constraints. The trouble is, when we transcribed English with additional complex segments (e.g., we transcribed *gumbo* with a prenasalized stop [mb]), still more improvement resulted, as shown in (20). The UCLA Phonotactic Learner achieved the best fit on the version of English with the segments in (20f), in addition to the usual singleton segments. Each of the six datasets in (20) assumes a different inventory of complex segments. Simulation results are ordered from worst fit to best (by Generality); there is a clear positive correlation between the number of complex segments assumed and the grammar's fit.

(20) Results for English phonotactic simulations assuming different complex segment inventories

	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

There are several reasons for this bizarre result. One is due to a design feature of the UCLA Phonotactic Learner: it does better when words are shorter (Daland 2015). Rewriting clusters as single segments reduces the average number of segments per word, so the learner has an easier time matching the attested distributions. This computational gain comes at a cost, since adding complex segments increases the number of natural classes and therefore constraints for the learner to sort through, but as long as that limitation can be overcome, shorter words automatically improve the learner's chances of fitting the data.

Another, more phonologically interesting reason is that rewriting certain clusters as segments allows the learner to capture gaps in the data in a very general way. To give a straightforward example, recall Ngbaka, whose complex segment inventory is [mb, nd, nz, ŋg, ŋm, ŋmgb, kp, gb, nw, vw]. If the learner sees these consonant sequences as segments, it can posit a very general constraint, *[-syll][-syll], as other clusters are marginal. If the learner sees them as clusters, however, it must induce a much larger set of constraints to characterize limitations on the set of acceptable clusters. These constraints include *[-son][+COR], *[-son,+LAB][-syll], *[-syll,+cont][+cons] (no continuant-initial clusters except [v w]), as well as a host of others. Thus, complex segments allow the learner to characterize restrictions on clustering with more generality.

Though the set of possible clusters in English (for example) is much larger, similar logic factors into the results in (20). If the English learner is exposed to a corpus where homorganic NCs are represented as prenasalized stops, for example, it can posit a very general constraint, *[+nas][-son], to explain why the heterorganic ones are rare. Without this option, it would have to posit a set of constraints banning each type of heterorganic NC (e.g. [+nas,+LAB][-son,+DOR], and so on; see Wilson and Gallagher 2018). And because the UCLA Phonotactic Learner 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 (20f) includes a constraint against prenasalized affricates, since those are rare in the language. So in fact there is a kind of perversity in this feature of the phonotactic approach. The more rare and restricted the complex segment, the better the fit.

Notably, we find the same pattern Russian that we find in English: the more sequences treated as complex segments, the better the result. We argued above that the phonotactic arguments for affricates in English and Russian are not as convincing as the statistical evidence supplied by the languages. The fact that the phonotactic learner performs best in precisely the situations where the distributional evidence for complex segments is lacking casts a serious doubt on phonotactic reasoning as a learning theory of complex segments.

5.3 Learning complexity from phonetics

It has been often hypothesized that complex segments differ phonetically from same-phone clusters—especially in duration (Trubetzkoy 1939 et seq.). Sagey (1986:79) notes that "if contour and complex segments are phonologically associated with single timing units [...] we would expect them to have the phonetic length of single consonants, rather than the length of consonant clusters, which occupy two timing units". Herbert (1986:10) defines prenasalized consonants as "exhibit(ing) the approximate surface duration of 'simple' consonants in those language systems within which they function." If this link between segmental status and duration is universal, then we might expect learners to exploit it: the learner might have a bias to unify short consonant sequences, in addition to (or instead of) inseparable ones. But while this correlation between segmental status and duration is often hypothesized, it has received little convincing support. In our view, the existing phonetic research raises more questions than it answers.

The following subsections discuss two reasons why we doubt that a universal correlation between segmenthood and duration exists, or that it could offer a general approach for learning the cluster/complex segment distinction. First, there are clear counterexamples (§5.3.1). Second, the inherent duration of segments and clusters can differ quite drastically both within and across languages; there is no principled way in many cases to decide what durations to compare (§5.3.2). Additional phonetic properties that could potentially differentiate segments from clusters are briefly discussed in §5.3.3.

5.3.1 Counterexamples

While some report a correlation between duration and segmenthood (Brooks 1964; Riehl 2008; Cohn and Riehl 2016), 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. Additional evidence comes from vowel reduction in closed syllables. Example (21) shows that [i, u, ɔ] appear in open syllables, and [ɪ, ʊ, a] in closed ones (cf. [p^{ɦ}ʊ̲k.ti̲] and [p^{ɦ}u̲.kɪ̲t]). The examples in (22) show that NC sequences behave like single segments: they are preceded by [i, u], just like [t] and [k] in (21) and unlike [k t] and [r n].

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

a.	[titɪp]	'leg'	cf.	a'.	[titi]	'meticulous'
b.	[kukʊr]	'scratch'		b'.	[kuku]	'finger'
c.	[p^{ɦ}ʊkti]	'evidence'		c'.	[p^{ɦ}ukɪt]	'hill'
d.	[sɪrnɔ]	'disappear'		d'.	[siram]	'bathe'

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

a.	[tiŋk^{ɦ}i]	'louse'	c.	[liŋg^{ɦ}ɪs]	'machete'
b.	[tuŋk^{ɦ}u]	'wait'	d.	[muŋkʊr]	'face down'

An analyst would have a good case for treating Javanese NCs as complex segments. Yet they are significantly longer than singleton stops and nasals in Javanese (Adisasmito-Smith 2004:307). Of course, it is possible that the evidence from distribution and alternations is misleading, and the sequences are represented as clusters, as Adisasmito-Smith (2004) ultimately claims. But the link between segmenthood and duration in this case is at best tenuous. The phonological arguments for segmenthood in Javanese are straightforward, but the durational data do not match them.

A second case of mismatch between duration and segmenthood comes from languages in the Bura-Margi cluster. They 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 the segmental treatments of Hoffman (1963) and 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, though he collects no data from this language) 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, too, appear to counterexemplify the claim that there is a link between duration and segmenthood: they pattern like single segments, yet are longer than single segments.

These counterexamples suggest that there is no universal link between duration and segmenthood. Of course, one could claim that the phonotactic evidence in these cases is misleading, and that duration correctly diagnoses these sequences as clusters. This is the strategy taken by Maddieson (1983:289) for Bura-Margi, Adisasmito-Smith (2004:313) for Javanese, and Riehl (2008:82) for Pamona. This move strikes us as circular: any research program which attempts to establish a connection between segmenthood and duration cannot use duration as an diagnostic for segmenthood without first establishing that there is a link between them. As languages that contrast complex segments with same-phone clusters are at best rare (Maddieson and Ladefoged 1993, Riehl 2008, and others) such a link has proven difficult to establish, likely in large part because there is no field-wide consensus for how to distinguish a complex segment from a cluster in the first place (Herbert 1986: Ch. 2).

5.3.2 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 1977:848). Nasals are considerably longer than stops in Sukuma (Maddieson and 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 and 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 and 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 explicitly in the literature on Modern Greek [t s] and [d z]. The analysis of these sequences has been the subject of much debate, with evidence from phonotactics, morphophonology, and phonetics recruited in favor of opposing analyses (Joseph and Philippaki-Warburton 1987; Tzakosta and 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 and 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 thus rejects phonetic arguments for or against affricate treatment. Neither phonetics nor phonotactics provide clear evidence to learners of Greek.

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 20 (the learner identified a total of 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 20: Inseparability measures for Modern Greek

5.3.3 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,[17] but the lack of a clear correlation between duration and phonological patterning casts 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 and 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).

More broadly, 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) 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 and 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 phonetic properties, as of yet undiscovered, that can reliably distinguish complex segments from clusters. In particular, it is an open question whether complex segments can be differentiated from same-phone clusters by their gestural organization. Saltzman and Munhall (1989), Löfqvist (1991), Byrd (1996) and others hypothesize that a segment is a constellation of gestures with a stable timing pattern. As noted by Byrd (1996:160), this definition of the segment allows us to make predictions about differences between complex segments and clusters. For NCs, for example, the specific prediction is that 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. Such a correlation would suggest the existence of reliable phonetic differences among NCs that could be directly linked to

[17] Brooks (1964) shows that duration is a cue to the cluster/affricate distinction in Polish [t̪ ʂ] vs. [ʈʂ], for example. The distinction is sometimes erroneously described as a same-place contrast (Clements and Keyser 1983:35), 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 [ʈʂ] is unified whereas [t̪ ʂ] is not (something we have verified in a simulation on Polish; again, results in Gouskova (in preparation).

a difference in their representational status. To our knowledge, there has been no work demonstrating reliable differences of this sort.

6 Experimental investigations of phonological awareness of complex segments vs. clusters

We have mostly measured the success of our learner against phonotactic argumentation, but the question arises whether the segment-cluster distinction can be tapped experimentally. This would perhaps be a more satisfying way of distinguishing segments from clusters, as it removes potential sources of bias on the linguist's part (see §3). But since phonological representations are covert, this is a difficult area to study in the lab. The main problem is that most studies of complex representations use metalinguistic experimental tasks that try to probe phonemic awareness. These come with seemingly simple linking hypotheses, but the results are not always interpretable as phonological knowledge.

6.1 Orthographic confounds

Several behavioral experiments have attempted to divine the difference between clusters and complex segments by asking people to insert vowels at locations of their choice (e.g. *bantsa* → *banitsa*, *bantisa*) or by breaking words up into syllables (*bant.sa* or *ban.tsa*). Skeptics point out that these experiments reveal not the complex segment/cluster representations but features of the orthography of the language (Arvaniti 2007), or the restrictions on possible phonological words rather than syllabification (Steriade 1999, Downing 2005). We share these concerns.

Orthography is clearly instrumental in attracting the attention of linguists to cases such as Greek [t s, d z], which are written as clusters, vs. [p s, k s], written with single letters. This feature of Greek orthography makes it difficult to interpret the results of metalinguistic experiments. For example, Tzakosta and Vis (2007) report[18] that Greek listeners were on average more reluctant to insert vowels into fricative-stop clusters than stop-fricative ones: epenthesis was most common in [d z, k s, t s, p s, s t] and less common in [s k, s p]. These results are uninterpretable given any reasonable hypothesis concerning Greek's consonant inventory, but they make sense if both the shorter duration of hormorganic clusters (affecting [t s, s t, d z]) and representation as a single letter (affecting [k s, p s]) affects epenthesis rates.

A closer look at the experiments probing the segment-cluster distinction shows that orthography is a confound even in languages where complex segments are not orthographically distinguished from clusters in a reliable way, such as English. In English, [tʃ, dʒ] are written in a variety of ways (*church, match, gin, jam, badge*). Some single segments are written as digraphs (*shoe*, *thin*, *phone*), and some orthographic clusters include silent letters (*know, gnome*, *psych*, *czar*, etc.). These orthographic features of English introduce confounds whenever phonemic awareness is required to peform a task, as is the case in Pig Latin. Pig Latin is a segment manipulation game, which in the simplest cases removes the first consonant of the word and suffixes it to the end followed by [ɛɪ], as in "cow" [kaʊ] → [aʊ-keɪ]. The game has two dialects: the "C1" dialect ("smile" [smaɪl] → [maɪl-seɪ]) and the "onset cluster" dialect ([smaɪl] → [aɪl-smeɪ]). Barlow (2001) used Pig Latin to study the structure of Cj clusters and sCC clusters, whose phonological status is debated. The question is whether speakers of "C1" Pig Latin will split sequences such as [kj] in "cute" or treat the [j] as part of the syllable nucleus. An anonymous reviewer suggests that this manipulation could also be used to study the status of [ts] and [tʃ]. English speakers should not be able to split [tʃ] if it is really a single segment, but [ts] should vary as other clusters do:

[18] We would not want to overinterpret these results, since the authors do not report statistical analyses and the plots do not show extent of variation in the results, just raw averages over 20 participants. In particular, the rates of splitting [k s, t s, p s, d z, s t] all look visually similar enough that they might not be statistically significant.

(23) Expectations for Pig Latin when tested on affricates vs. clusters

orthography	IPA	in Pig Latin "onset" dialect	in Pig Latin C1 dialect
scout	skaʊt	aʊt-skeɪ	kaut-seɪ
church	tʃɚtʃ	ɚtʃ-tʃeɪ	ɚtʃ-tʃeɪ (*ʃɚtʃ-teɪ impossible!)
tsunami	sunɑmi	unɑmi seɪ	unɑmi-seɪ
	tsunɑmi	unɑmi tseɪ	sunɑmi-teɪ

As Barlow found, however, several of her participants did indeed split "church", though not along the phonological lines. Several speakers split orthographic <h> from "ch", so "church" was [hɚtʃ-tʃeɪ], and "chew" was [hju-seɪ] and [jutʃeɪ] (her Appendix A and B). One even split "shoe" as [hu-seɪ], and "thumb" as [hʌm-θeɪ]. This is not a phonological pattern but an orthographic one. Barlow suggests that orthographic confounds could be avoided by testing preliterate children, but we are dubious. Research on reading proficiency demonstrates that Pig Latin proficiency is strongly correlated with reading ability, a special kind of metalinguistic awareness that is developed through explicit training. In a large-scale study of various correlates of reading ability (using 65 third-graders), Hester and Hodson (2004) found that Pig Latin is the strongest correlate of ability to read unfamiliar words. Similarly, in a longitudinal study of Pig Latin acquisition on a boy aged 5 to 7, Cowan (1989) found that the boy struggled with the task at age 5 (only 14.3% to 47.1% correct, depending on the length of the target word), and his accuracy improved to 30-92% only when his reading skills developed. This strongly suggests that segment manipulation tasks cannot be divorced from orthographic knowledge, and that Barlow's hypothesis about preliterate children being a better population for this sort of study cannot be tested. Children cannot do the task.

6.2 Metalinguistic and non-metalinguistic tasks: Quebecois French affricates

We suspect that more convincing evidence for the single-segment status of complex segments will come from experimental paradigms that use non-metalinguistic tasks to probe speakers' knowledge more subtly. Studies of this sort are, however, uncommon. One exception we know of is Béland and Kolinsky's (2005) study of Quebecois and European French. These dialects differ with respect to an assibilation rule, not reflected in the orthography. In Quebecois French, the rule turns the dental stops {t, d} into {ts, dz} before high front vowels {i, y}, such that <tu> /t y/ 'you' is pronounced [ts y]. In European varieties of French, no assibilation occurs (Walker 1984, Béland and Kolinsky 2005).The goal of their study was to determine if a nonce word repetition task (non-metalinguistic) and a backwards talking task (metalinguistic) could shed light on speakers' representations of these sequences.

The repetition experiment asked people to repeat nonce words produced by a Quebecois talker. Quebecois listeners repeated {ts, dz}/{i,y} with affricates most of the time, whereas Belgian listeners produced {t, d} in those contexts most often, {ts, dz} some of the time, and occasionally [θ]. For non-allophonic contexts such as [etsa] and [usto], both groups repeated the words accurately most of the time. This result is important: it suggests that the stimuli produced by the Quebecois speaker had acoustic differences in the duration of frication in [ts, dz] followed by {i,y} vs. [ts, dz] followed by other vowels. This raises the question of whether the European listeners could even hear {ts, dz}{y/i} as having frication. Béland and Kolinsky analyze their stimuli acoustically and find that the allophonic [ts, dz] sequences (before {i, y}) are shorter than [ts, dz] before other vowels, supporting this interperation.

In the backwards talking task, participants were trained to say [ifa] as [afi]. They were not exposed to written stimuli. The hope is that two consonants would be reversed if they form a cluster, /a t s i/ → [ista], but not if they form a complex segment, /a ts i/ → [itsa]. Since it is doubtful that European French listeners perceived [ts], [dz] as affricates before {i, y}, it is difficult to interpet the results of this task. Unsurprisingly, Quebecois speakers reversed [itsyfa] as [afytsi]; this is expected since [ts] is required both before [y] and before [i]. Also unsurpising (given the repetition task) is that European French speakers reversed [ts] to [t] most of the time, and that they also reversed [ts] to [st] more often than Quebecois speakers did. This might be taken to mean that when European speakers heard [ts] as [ts], they treated it as a sequence of two segments. But what do we make of the finding that Quebecois speakers also sometimes reversed [ts] to [st], too? Do these speakers represent allophonic [ts] as a sequence of two segments, or is the metalinguistic nature of the task affecting their behavior? On the one hand, this experiment suggests that European French speakers might be treating Quebecois allophonic [ts] as two segments; on the other, it suggests that even some Quebecois speakers treat it as two.

Our simulations on French failed to find any complex segments when the transcriptions reflected European French phonology, but the results were ambiguous when we used Quebecois transcriptions. We tested Lexique, a 71,252-word list (lexique.org), as well as two versions of CHILDES child-directed speech corpora: 115,1143 words as connected speech, or 8,896 words when tokenized. In the connected speech corpus, words such as [t y]/[t s y] 'you' appear multiple times; in the tokenized "type frequency" corpus, only once. We found that in all three simulations, [d z, t s] are at the top in terms of inseparability, but the sequences only pass the threshold for unification in the CHILDES CDS connected speech simulation (shown in Fig. 9). In European French, nothing comes close to complex segment status; its first iteration looks like the 2nd iteration for the Quebecois simulation in virtually every simulation.

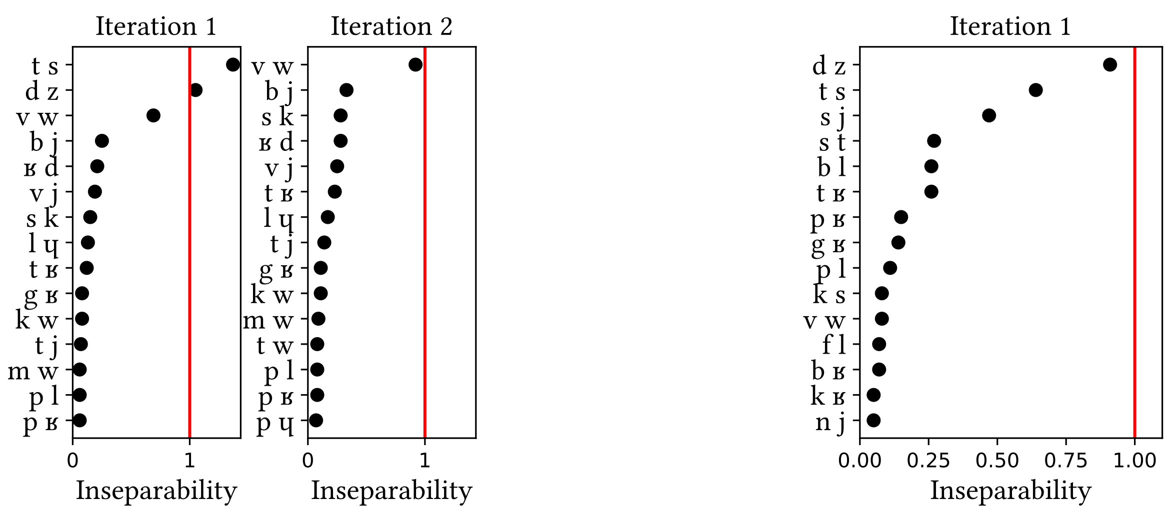

Figure 9: Quebecois French simulation:, connected child-directed speech (left) vs. Lexique (right)

This learning exercise could be taken to mean that Quebecois French supplies weak distributional evidence for its complex segments, resulting in inconsistent learning of complex segment representations. For some speakers, the sequences are clusters, and for others they are affricates. This is consistent with the experimental results described above.

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 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 supported by phonotactic argumentation (and in some cases experimental evidence) in in a large majority of the cases we have discussed.

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 representatoins. We also evaluated currently available evidence for phonetic differences between purported complex segments and clusters, and concluded that these differences are not consistent enough to be a reliable cue to distinct representations for learners. 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 also have a ways to go before we understand 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 phonological representations we posit are actually real.

References

Adisasmito-Smith, Niken. 2004. Phonetic and Phonological Influences of Javanese on Indonesian. Doctoral Dissertation, Cornell University, Ithaca, NY.

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.

Anderson, Stephen R. 1976. Nasal consonants and the internal structure of segments. *Language* 52:326–44.

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 Blake Allen. submitted. Learning alternations from surface forms with sublexical phonology. *Phonology*, URL `http://ling.auf.net/lingbuzz/002503`.

Becker, Michael, Nihan Ketrez, and Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish devoicing neutralization. *Language* 87:84–125.

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. Doctoral Dissertation, University of Massachusetts, Amherst.

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. URL `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 Elizabeth Hume. 1995. The internal organization of speech sounds. In *The Handbook of Phonological Theory*, ed. John A. Goldsmith, 245–306. Cambridge, MA, and Oxford, UK: Blackwell.

Clements, George N., and Samuel Jay Keyser. 1983. *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? In *Studies in language typology and change*, ed. Yanti and Timothy McKinnon, volume 60 of *NUSA*, 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. Doctoral Dissertation, Rutgers University, New Brunswick, NJ.

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. In *Handbook of Australian Languages*, ed. Robert M. W. Dixon and B. J. Blake, volume 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*. Harare: Mercury Press, 2nd edition edition.

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. 2010. Perceptual distinctness and long-distance laryngeal restrictions. *Phonology* 27:435–480.

Gallagher, Gillian. 2011. Acoustic and articulatory features in phonology–the case for [long VOT]. *The Linguistic Review* 28:281–313.

Gallagher, Gillian. 2013. Speaker awareness of non-local ejective phonotactics in Cochabamba Quechua. *Natural Language & Linguistic Theory* 31:1067–1099.

Gallagher, Gillian. 2016a. Asymmetries in the representation of categorical phonotactics. *Language* 92:557–590.

Gallagher, Gillian. 2016b. Vowel height allophony and dorsal place contrasts in Cochabamba Quechua. *Phonetica* 73:101–119.

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. 2018. Phonotactic restrictions and morphology in Aymara .

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.

Göksel, Aslı, and Celia Kerslake. 2004. *Turkish: A comprehensive grammar*. New York: Routledge.

Gordon, Matthew K. 2016. *Phonological Typology*. Number 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, 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. URL `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.

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). In *The Bantu Languages*, ed. 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. In *NELS 46: Proceedings of the Forty-Sixth Annual Meeting of the North East Linguistic Society*, ed. 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. In *21st Manchester Phonology Meeting*.

Ito, Junko. 1986. Syllable Theory in Prosodic Phonology. Doctoral Dissertation, University of Massachusetts, Amherst. 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. In *Ohio State University Working Papers in Lingiustics*, ed. Marivic Lesho, Bridget J. Smith, Kathryn Campbell-Kibler, and Peter Culicover, volume 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. Doctoral Dissertation, MIT, Cambridge, MA. 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. In *Phonology at Santa Cruz*, ed. Aaron F. Kaplan and David Teeple, volume 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*, volume 14. Lincom Europa.

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

Kochetov, Alexei. 2006. Syllable position effects and gestural organization: Articulatory evidence from Russian. In *Laboratory Phonology 8*, ed. 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.* Chicago: University of Chicago Press, 2 edition.

Ladefoged, Peter, and Ian Maddieson. 1996. *The sounds of the world's languages*, volume 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. In *Blackwell companion to phonology*, ed. Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice, volume 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. Doctoral Dissertation, UCLA, Los Angeles.

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. In *UCLA Working Papers in Phonetics*, volume 83, 1–7. UCLA Department of Linguistics.

Maddieson, Ian, and Peter Ladefoged. 1993. Partially nasal consonants. In *Nasals, Nasalization, and the Velum*, ed. 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. Doctoral Dissertation, UCLA, Los Angeles, CA.

McCullagh, Matthew. 2011. The Sounds of Latin: Phonology. In *A Companion to the Latin Language*, ed. James Clackson, 81–91. Hoboken, NJ: Blackwell Publishing Ltd.

Mohanan, K. P. 1982. Lexical Phonology. Doctoral Dissertation, MIT, Cambridge, MA. 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. Doctoral dissertation, University of British Columbia, Vancouver, BC.

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. In *Proceedings of the 3rd Annual Meeting of the Berkeley Linguistics Society*, ed. 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. In *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. In *The Prosody-Morphology Interface*, ed. René Kager, Harry van der Hulst, and Wim Zonneveld, 310–343. Cambridge: Cambridge University Press.

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. Doctoral dissertation, Ithaca, NY.

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. In *Studies in Linguistic Analysis*, ed. 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. Doctoral Dissertation, Université Grenoble, Grenoble.

Rubach, Jerzy. 1985. Affricates as strident stops in Polish. *Linguistic Inquiry* 25:119–143.

Rubach, Jerzy. 2000. Glide and glottal stop insertion in Slavic languages: A DOT analysis. *Linguistic Inquiry* 31:271–317.

Sagey, Elizabeth. 1986. The Representation of Features and Relations in Nonlinear Phonology. Doctoral Dissertation, MIT, Cambridge, MA. 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. In *The Structure of Phonological Representations*, ed. Harry van der Hulst and Norval Smith, volume Part II. Dordrecht: Foris Publications.

Shih, Stephanie S, and Sharon Inkelas. 2018. Autosegmental Aims in Surface-Optimizing Phonology. *Linguistic Inquiry* 50:137–196.

Shih, Stephanie S, and Sharon Inkelas. 2019. 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. In *IPO Annual Progress Report*, ed. A. J. Breimer, volume 9, 64–69. Eindhoven: Institute for Perception Research.

Stanton, Juliet. 2016. Predicting distributional restrictions on prenasalized stops. *Natural Language & Linguistic Theory* 32:1089–1133.

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

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. In *Nasals, nasalization, and the velum*, ed. Marie K. Huffman and Rena A. Krakow, 401–470. San Diego: Academic Press.

Steriade, Donca. 1999. Alternatives to syllable-based accounts of consonantal phonotactics. In *Proceedings of the 1998 Linguistics and Phonetics Conference*, ed. Osamu Fujimura, Brian Joseph, and B. Palek, 205–242. Prague: Karolinum 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. URL `http://www.speakrus.ru/dict/index.htm#tikhonov`.

Tlale, One. 2005. The phonetics and phonology of Sengwato, a dialect of Setswana. Doctoral Dissertation, Georgetown University, Washington, D.C.

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.

Walker, Douglas C. 1984. *The pronunciation of Canadian French*. Ottawa: University of Ottawa Press.

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. Doctoral Dissertation, UCLA, Los Angeles, CA.