Getting ready for primetime: paths to acquiring substance-free phonology[*]

Bridget D. Samuels[1], Samuel Andersson[2], Ollie Sayeed[3], & Bert Vaux[4]
[1]University of Southern California; [2]Yale University; [3]University of Pennsylvania;
[4]Cambridge University

To appear in the *Canadian Journal of Linguistics* special issue on features

Abstract
Substance-free phonology (SFP) is based on the hypothesis that phonological computation makes no reference to phonetic substance, and that phonological features are treated as arbitrary symbols for the purposes of computation. However, phonologists within the SFP tradition disagree about whether the content of phonological features is innate or learned ("emergent"), and if learned, whether the acquisition process is based on phonological patterning alone or refers to phonetic substance. In the present work we identify predictive differences between these accounts. We conclude that there is an innate basis to phonological features, but that featural content is not innate. We suggest that a hybrid phonetic-phonological approach to feature content acquisition may ultimately be the most successful.

Keywords features, acquisition, substance-free, innateness

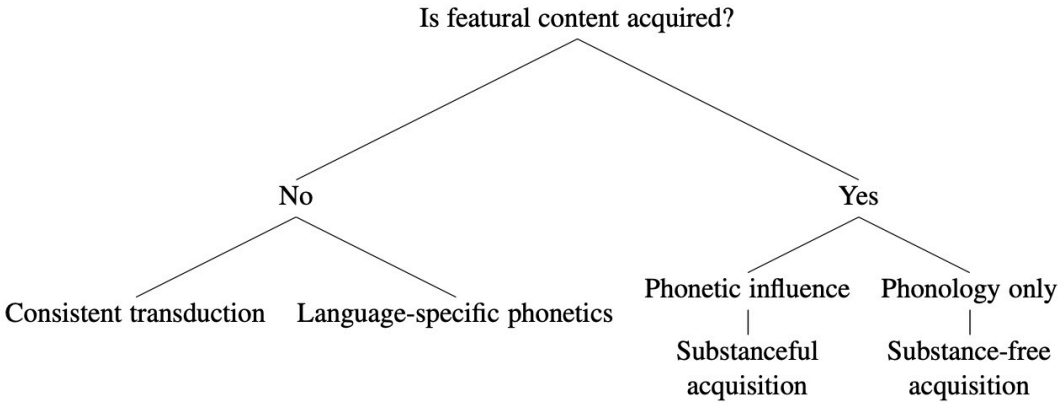
1. Introduction

In this work, we attempt to shed light on the question of *how much* of phonology is free of phonetic substance.[1] At its core, substance-free phonology (SFP) is based on the hypothesis that phonological computation makes no reference to phonetic substance; phonological entities like features are treated as arbitrary symbols for the purposes of computation (Hale & Reiss 2000: 162). There remains widespread disagreement concerning whether phonological primes (features), too, are substance-free in their acquisition and/or their ultimate acquired state. As we see it, there are four possibilities, outlined in (1):

[*] We would like to thank David Odden, Elan Dresher, Johanna Benz, the audiences at the Phonological Theory Agora in March 2019 and the Workshop on Theoretical Phonology in May 2020, as well as two anonymous reviewers for their helpful comments on previous versions of this material.

[1] Following Hale & Reiss (2000), the notion of phonological 'substance' is opposed to that of phonological 'form.' Specifically, the phonetic correlates of a particular feature are its substance, and the core hypothesis of substance-free phonology holds that "[p]honological primes are substance-free, in that their phonetic interpretation… does not play a role in phonological computation" (Blaho 2008: 2). Asymmetries that may be observed in the way particular features behave are attributed to extra-phonological properties, e.g. articulatory or perceptual biases, rather than markedness statements within the phonological grammar.

(1) Feature acquisition decision tree



The traditional view in substanceful phonology, also held by some practitioners of SFP (e.g. Hale & Reiss 2008 et seq.), takes both the concept of phonological features and their content to be innate. On this view, children are born with a set of phonological features that map to certain acoustic and/or articulatory targets, and phonological acquisition involves learning which of these features are active in the child's target language(s). Innate features may be consistently transduced into phonetic content (see Volenec & Reiss 2017 for a proposal). Alternatively, phonetics may be at least partly unpredictable on a language-specific basis, and therefore acquired (e.g. Keating 1984). We discuss these possibilities in Section 2.

Adopting SFP raises the possibility that the content of features may be acquired based on either phonetics, giving us sets of segments that share articulatory or acoustic properties, or phonology, giving us sets of segments that pattern together in the language. This has sometimes been called the "emergentist" view (e.g. Mielke 2008). Acquiring featural content is a logical possibility only if the SFP hypothesis is correct, and no innate components of phonological computations refer to phonetic content; this approach is also motivated by a more generally minimalist view of the language faculty that attempts to eliminate redundancy while taking into account evolutionary plausibility (on implications for phonology, see e.g. Samuels 2011). Some works have posited that features are acquired purely on the basis of phonological patterning (e.g. Odden this volume, Mayer & Daland 2019); language-specific features constructed in such a manner are themselves substance-free and may not be transducible into coherently describable phonetic content. We call this the "phonological feature-learning" view. On the other hand, acquisition of features may be at least partly determined by their phonetic properties, such that features are ultimately transduced to phonetic content, though phonological computation may not be able to refer to that content. We call this the "phonetic feature-learning" view; see Lin (2005) and Mielke (2012) for attempts to identify features via unsupervised clustering on acoustic and articulatory data. We discuss the phonological and phonetic feature-learning

views in 3.1 and 3.2, respectively.

In the present work we point to a number of predictive differences between these accounts. We carefully consider both the philosophical and empirical arguments for and against the innatist view, focusing on data that could help decide between--or bring together--phonetic and phonological feature-learning accounts. We conclude that there is an innate basis to phonological features (including, at the very least, the concept of a feature), but that the content of features is not innate. Phonetic and phonological feature-learning approaches have different shortcomings, collectively and individually, and will need revision and augmentation to account for the range of attested data. We suggest that a feature-learning approach that takes phonetic information into account but allows phonology to overrule the phonetics may ultimately be the most successful.

2. The innatist view

Although a range of work has been done under the name 'substance-free phonology', there is disagreement within the substance-free literature concerning where substance does and doesn't appear in the process of acquiring a language. As noted in (1), to a first approximation, there are two schools of thought regarding the origin of phonological features: either they are innate, or they are learned. Innatists in the substance-free tradition could in principle permit language-specific phonetics, although we are not aware of any practitioners of SFP who do so.

The innatist position within SFP is perhaps best represented by the work of Mark Hale and Charles Reiss, who have argued for "the logical necessity of a discrete, innately available system for phonological representation" (Hale and Reiss 2008: ix). In later work by these authors and others, the mapping from phonological features to articulatory instructions is also taken to be innate and invariant across languages, so that there is no language-specific phonetics (Hale, Kissock and Reiss 2007; Hale and Reiss 2008: 116-117; Volenec and Reiss 2017, this volume). The arguments for innate features and universal phonetics are explored in the rest of this section.

2.1 Card grammars and phonological grammars

The argument that innate features are a 'logical necessity' is illustrated in Hale and Reiss (2003) by using toy languages involving playing cards. Consider a learner with only the innate feature [DIAMONDS], present for all and only diamonds, and otherwise absent. Below in (2) are some input cards, and the way that a learner with this impoverished feature system would parse them:

(2) Toy card grammar

Input	Parse by learner
5 of diamonds	[DIAMONDS]
King of diamonds	[DIAMONDS]
7 of hearts	---
7 of spades	---

Since the only feature available encodes suit, there is no representation of a card's value: both the 5 and king of diamonds are assigned the same representation. Non-diamond cards are not parsed as linguistic information at all; Hale and Reiss (2003, 2008) draw a comparison to the parse of a belch by the phonological system. Since many of these contrasts are not parsed, they can never be acquired, the argument goes. The learner can never begin to notice that the 5 and king of diamonds are distinct cards, because their representational system provides no way of distinguishing them. Similarly, the learner can never acquire a distinction between hearts, spades, and clubs because noticing this difference is beyond the power of the learner's representations. Because of this, with card grammars as well as phonological grammars, all possible contrasts that languages can make must be available innately to the representational system.

Hale and Reiss (2008: 116-117) point to a learnability problem for the idea that the phonetics-phonology interface may vary across languages if features are innate. Without knowing which feature a particular acoustic output comes from, there is no way of knowing what the surface form, or output of phonological computation, might be. Perhaps a vowel like mid-centralized ɪ̈ is really targeted as [i] (with language-specific lowering and centralization) or [e] (with language-specific centralization), or [ɨ] (with language-specific lowering), and so on. Without knowing what the surface forms are, there is no way to set up an underlying form or any phonological processes, and so phonological learning cannot progress. Taking this view to its extreme leads to the rejection of any language-specific phonetics. This entails that consistent phonetic differences between languages must be represented phonologically:

> "In our view, for example, the recurrent difference in pronunciation of English [i] and German [i] is to be attributed to representational (featural) differences present in I-languages of English and German speakers, *not* to language-specific phonetics. In general, our position is that all recurrent or linguistically relevant differences in pronunciation result from representational differences in the lexicon and from differences in the phonological rule component." (Volenec and Reiss 2017: 272)

Having summarized two important parts of the innatist view from SFP - innate representational primes, and universal phonetics - we now discuss several arguments against this position. In Section 3 we present alternatives.

2.2 The card grammar argument does not preclude feature learning

We propose that phonological features build on innate infrastructure, but that their content is not itself innate. In other words, we agree with the position set forth in 2.1 that innate representational primitives are necessary for learners to begin learning a phonological system, but we contend that these primitives are not phonological features. The possibility that there are "more basic primitives at the initial state" is raised by Hale & Reiss (2008:37), who conclude that these initial primitives would still nevertheless be part of UG. Converging lines of evidence instead suggest to us that the initial primitives are (a) properties of the auditory system that are innate, but not exclusively

phonological[2], and (b) an evolutionary inheritance of (at least) the mammalian lineage, shared with animals that do not have phonology (see also Samuels 2011, 2012).

One of the strongest cases for an innate but pre-phonological psychoacoustic bias influencing phonological systems is the +20 ms voice onset time (VOT) boundary utilized by many languages. This is known as the positive auditory discontinuity, and represents a non-linear mapping between the acoustic input and the associated percept (Kuhl & Miller 1975; Keating 1984; see Holt et al., 2004 for an overview). We are particularly sensitive to contrasts in both speech and non-speech stimuli that straddle auditory discontinuities; in other words, these psychoacoustic biases enhance our categorical perception. As Holt et al. (2004: 1763) put it, "[l]anguages may capitalize on regions of perceptual space where sensitivity is enhanced, adopting sounds for which moderate changes in articulatory or acoustic characteristics result in disproportionately large perceptual consequences." The positive auditory discontinuity has been confirmed by many perceptual studies of VOT and its non-speech analogue, tone-onset time (TOT), which can be defined as the temporal difference between the onset of low-frequency energy and the onset of higher-frequency energy (Hay 2005). Even humans whose native languages do not include category boundaries at these discontinuities remain particularly sensitive to them (Williams 1974, Streeter 1976, Hay 2005). In experiments testing TOT discrimination, it was found that distributional boundaries coinciding with the +20 ms and -20 ms auditory discontinuities appear easier to learn than boundaries at +40 ms and -2.5 ms TOT, which do not coincide with discontinuities (Holt et al., 2004). Enhanced sensitivity to these TOT values is already detectable in 2.5-month-old infants (Jusczyk et al., 1980).

The significance of these results for the question at hand is difficult to interpret: there are no truly 'pre-linguistic' infants, and the relationship between speech and non-speech auditory processing is not entirely clear.[3] For this reason, it's important to look to other species. Since the 1970s (e.g. Kuhl & Miller 1975, Waters & Wilson 1976, Kuhl & Padden 1982, Kluender et al. 1987, Dooling et al. 1989), studies on animals' ability to perceive and categorize auditory stimuli, including their own acoustic communications and those of humans, have brought substantial evidence to bear on the question of whether speech is special in the way it is processed by our species. In their review of thirty years of literature in this area, Brown & Sinnott (2006:198) concluded that "humans are not much more sensitive than other animals to differences between speech sounds" and perceived similar boundaries for 17 of 27 tested contrasts including *ba-pa*, *ra-la*, and *ba-wa*. Neuroimaging and psychoacoustic studies have shed further light on the temporal dynamics of auditory processing. Comparative studies of the acoustics of animal calls and human speech led Suga (1969, 1973) to conclude that there are at least three basic acoustic elements shared across species--constant-frequency components,

[2] We set aside other sensory modalities for the time being. The same arguments should hold for, e.g., visual primitives of signed languages; see 3.3.

[3] It has commonly been assumed that speech and non-speech processing are processed similarly but independently (Liberman et al., 1967; Liberman & Mattingly, 1985), but see, e.g., the findings of Bent et al. (2006) and Berent et al. (2010) that linguistic experience appears to affect performance on non-linguistic auditory tasks.

frequency-modulated components, and noise-burst components--and to hypothesize that these elements were recognized by specialized types of neurons. Distinct clusters of neurons in auditory cortex that respond to combinations of these properties have since been revealed in amphibians, birds, bats, and primates (see Suga 2006 and references therein). The picture that emerged is one in which "the basic principles operating for processing species-specific complex sounds in amphibians, avians, and nonhuman mammals are greatly shared with the human auditory system for processing 'speech sounds,' and the human auditory system has developed highly specialized mechanisms for processing 'speech' from shared mechanisms" (Suga 2006:177). Suga's hypothesis is further supported by recent neurophysiological studies, such as that of Mesgarani et al. (2008), who found that individual neurons in ferret auditory cortex are sensitive to important acoustic properties of human speech, such as specific frequency bands, formant transitions, broad-spectrum noise (frication), and rapid transitions between silence and noise (e.g. plosive release bursts). Phonological features may indeed find their basis in these shared sensitivities, but this does not entail that the features themselves are innate as such; these sensitivities are not themselves features, either (a point also made by Dresher 2009, 2014).

Taken together, these studies strongly suggest that phonological features and categories build on the sensitivities of an auditory system that is largely shared with other mammals. In Oudeyer's (2006: 53) terms, this stance is nativist in a "morphoperceptual" sense: the neurophysiological and psychoacoustic building blocks of phonology are innate, but they evolutionarily pre-date human language.

The heart of the matter therefore lies in what human specialization enables infants to go from categorizing sounds with the help of these sensitivities to having a full-fledged system of phonological features. This remains somewhat mysterious, but it likely involves warping the cortical perceptual map specifically for speech sounds, while maintaining access to the 'un-warped' map when perceiving non-speech (see Samuels 2012 for a more concrete proposal). We concur with Odden (this volume), and indeed with Hale & Reiss (2008), that in addition to the properties of the auditory system outlined above, humans must be innately endowed with the formal concept of a phonological feature, as well as a syntax of features and the phonological computations over which they operate. It is only by virtue of having such innate architecture that we can make crosslinguistic predictions about what phonological processes look like (e.g. A → B / C_D, where A-D are feature bundles), or how phonological processes interact (e.g. through extrinsic ordering). To the extent that there are universals in these domains (see Vaux 2008 and Andersson in press for typological advantages of serial rule-based theories), only an innate architecture can capture them.

We depart from Hale & Reiss, however, in contending that the language acquisition device must then acquire both the features and the language-specific nature of the computations. Although it may seem like a daunting task to construct a feature system, it is worth noting that we can construct new categories from highly variant acoustic input in non-linguistic situations. Think of Beethoven's 5th symphony, opening with its famous four-note motif -- absent obvious pathology, we can recognize this piece played by different instruments, or in a different studio with different acoustics, or in a

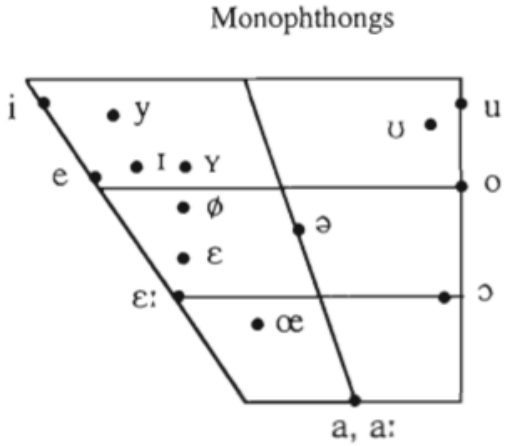
disco style, or in a different key, or at a different tempo, or on an untuned instrument, or with errors, and so on.

2.3 Consequences of innate transduction of features

In 2.1 we saw that the innatist view has been extended to transduction, so that the mapping between phonological feature bundles and articulatory instructions is taken to be invariant across languages. Volenec and Reiss (2017: 272) describe their position as "somewhat controversial", and it has many consequences, especially when paired with a computational system which relies on "abstract, symbolic, discrete, timeless units" (Volenec and Reiss 2020: 48). This section explores some of these consequences, and highlights areas where additional work is necessary. Odden (this volume) provides further insights into the cross-linguistic landscape.

As Volenec and Reiss (2017) note, in their theory there is necessarily a featural difference between German [i] and English [i], since there are consistent phonetic differences between the pronunciations of these surface forms in the two languages. However, they note that with 3 feature values (+, -, and unspecified) for 20 different features, we get 3^{20}, or roughly 3.5 billion, distinct segments. Even with a small number of features, then, we can describe large numbers of segments. However, this ignores the fact that the features and their transductions are meant to be innate. Consider [i] within the phonological system of German (3):

(3) German vowel system (reproduced from Kohler 1999: 87)



[i] is surrounded on almost all sides by other contrastive vowels, and so is presumably [+high] (to be distinct from [e]), [+ATR] (to be distinct from [ɪ]), and [-round] (to be distinct from [y]).[4] Although there is no [ɨ] in German, it is clear from the chart above

[4] Some of these feature values may be underspecified: [0high, 0ATR, 0round] etc. We use + since there is no phonetic evidence for such underspecification in German, as there is in e.g. Marshallese (Bender 1968, Choi 1992, Hale 2000). Even if we allow 0 and + options for all of these features across German-like I-languages, this would only give us 8 different options (three features with two possible values each).

that [i] is the most front vowel in the entire inventory, and so we treat it as [-back]. We disagree with Volenec and Reiss (2020: 55) who say that "there are many different feature sets (segments) that inhabit the 'high front unrounded vowel' space." We have now used up all of the common features used to distinguish plain (non-nasal, non-creaky, …) vowels, and the result is only a single [i]. But there are millions of speakers of German! Accounting for minute dialectal variations from one town to the next would require at least several thousand [i]-like vowels, but a small (conventional) feature set gives us only one. What if we were to add all of the possible [i]-like vowels found across all speakers of all varieties of Germanic? We would likely need several million possible [i]-like vowels, all of which are representationally identical in current systems with 20-30 features.

The problem seen with German(ic) [i] can be extended to any phonetic contrast (see also Keating 1984 for related arguments): a theory with invariant transduction would have to represent thousands of different laryngeal settings, thousands of ways of grooving one's tongue for a sibilant, and so on. The observation that 3^{20} is a large number is largely irrelevant, since each segment is only represented by a tiny number of feature bundles within this large space. If we are to capture every *possible* segment as realized in every *possible* human language, we would need to be able to represent thousands of variants of each IPA symbol, unlike current theories where each IPA symbol has at most a handful of possible feature representations. If it is possible to do this with only a small number of features, it should be easy to write down the complete list of these features, along with the innate transductions of each feature to articulatory instructions. Such a proposal would then make concrete, falsifiable predictions, which could be tested by analyzing large sets of phonetic data on similar speech sounds across different languages.

We must also remember that much of the featural variation discussed above exists within individual speakers. We consider one example based on the Dutch rhotic below, but it must be emphasized that virtually any study of intraspeaker variation could have been used. Vieregge and Broeders (1993) report on Dutch speakers' realizations of the rhotic phoneme in different phonological environments. Their Speaker 5 uses at least 12 different realizations of the rhotic, and all of these can be found within a single phonological environment (the coda). Their I-language would therefore need at least eleven optional rules to derive the eleven coda realizations which are not the underlying form.

Speaker 5's twelve realizations were all found in a sample of only 53 coda tokens. How many variants would exist if we considered 530, or 5,300 tokens?[5] And what if we considered all Dutch phonemes rather than just one, in all phonological environments? How many optional rules would be needed for a full grammar of Speaker 5? Given the range of intraspeaker variation, especially when fine phonetic detail is considered, it seems likely that the vast majority of rules in the vast majority of grammars would be

[5] Of course the realizations studied in acoustic work are bodily outputs rather than surface forms or articulatory instructions, such that some of the variation observed is not phonological. However, in the case of Dutch rhotics, it should be clear that variants like [r, ʁ, j, ɻ, ɛ, ø] (Sebregts 2015: 281) require distinct sets of articulatory instructions.

optional.

Despite the huge importance of optionality under the innatist view, the current innatist substance-free literature talks only about categorical and obligatory rules manipulating discrete symbols. There are many possible analyses of optionality, each with empirical consequences (see Anttila 2007 for an overview). For example, we could let features vary according to independent optional binary parameters, which makes strong predictions about how optional properties combine. For example, for all Dutch speakers who produce [r] (coronal trill), [ʀ] (uvular trill), and [ɹ] (coronal approximant), optional binary features predict that uvularity and approximance can also be combined to give [ʁ̞]. A speaker with only [r], [ʀ], and [ɹ] is predicted to be impossible. Is it true that no such speakers exist? What do other theories of optionality predict? Which approaches are empirically preferable to capture the many cases of intraspeaker variation from the previous decades of sociophonetic research? Carrying out these empirical studies on variation in phonetic detail strikes us as one of the key priorities for the innatist view, and we hope such work receives the attention it deserves in future literature from this perspective.

2.4 Evidence from non-speech modalities

In 2.4.1 we explore a case study from Gomera Spanish and its whistled equivalent, Silbo Gomero, in some detail. Our central argument is that there is only a single phonological grammar underlying both the spoken and whistled versions of this language. This means that some (I-)language-specific mapping must happen after the output of the phonology, to derive the divergent bodily outputs of spoken and whistled Spanish. We also show that the same phonetic contrasts can be put to different phonological use in speech and whistling, which we take to challenge the idea that there is a one-to-one mapping between phonetic properties and phonological features. The same argument could be made by considering any other whistled languages, or by analyzing visual and tactile versions of sign languages. We also discuss some of the challenges raised by sign languages for an innate feature system in 2.4.2.

2.4.1 Whistled language

Gomera Spanish is the variety of Spanish spoken on La Gomera, one of the Canary Islands. On this small and mountainous island, a whistled language called Silbo Gomero is used to convey messages over longer distances than are possible with speech. Gomera Spanish and Silbo Gomero are not, however, different languages: it has been clear since the 19th century (Lajard 1891 [1976]) that they have the exact same lexicon, morphology, syntax, and so on, differing only in how the outputs of the grammar are realized. Moreover, the functional neuroimaging study by Carreiras et al. (2005) revealed that people proficient in Silbo Gomero, but not Spanish-speaking control subjects, showed activation of language-related cortical regions during passive listening as well as during active monitoring tasks in both whistled and spoken forms. This suggests that for whistlers, Silbo Gomero is processed as language, and that the mapping between speech and whistling is not something metalinguistic added on after ‘true’ linguistic processing.

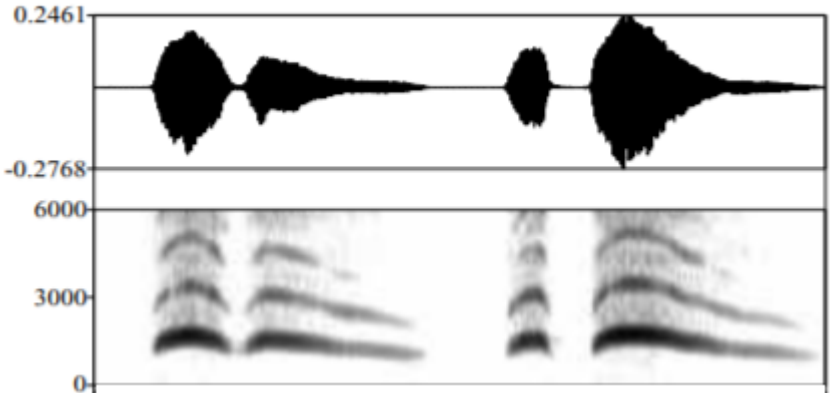
Below in (4) are the phonemes of Gomera Spanish, and the corresponding 'whistlemes' of Silbo Gomero, along with their acoustic realizations. Note that Silbo Gomero has fewer contrasts than Gomera Spanish (see e.g. Nevins 2015 on the reduction in vowel contrasts), although the set of contrastive units listed here is arguably too small, and additional distinctions may be made and sometimes perceived by some whistlers some of the time. We rely here on the minimal set of contrasts, which have high rates of recognition in perceptual experiments, and which are used in teaching of Silbo Gomero to local students (Classe 1957, Trujillo 1978, Meyer 2005, Rialland 2005).[6]

(4) Gomera Spanish and Silbo Gomero

Silbo Gomero	Gomera Spanish	Silbo Gomero acoustic correlates (our symbols)
/I/	/i, e/	Sustained higher pitch (roughly 2,000-2,500Hz)
/A/	/a, o, u/	Sustained lower pitch (roughly 1,500Hz)
/D/	/d, ɾ, r̲, l, n, ɲ, j, ʎ/	Rise-fall (whistled throughout)
/T/	/t, s, t͡ʃ/	Rise-fall (middle portion silent)
/B/	/b, m, g/	Fall-rise (whistled throughout)
/P/	/p, f, k, χ/	Fall-rise (middle portion silent)

The waveform and spectrogram below in (5), reproduced from Rialland (2003: 2133), show whistled /ABA/ on the left realized as a continuous fall-rise, and /APA/ on the right, with an interrupted fall-rise contour.

(5) Whistled /ABA/ and /APA/



Collapsing 23 phonemes into 6 whistlemes naturally results in extensive homophony. For example, whistled /PATA/ may correspond to spoken /pata, pasa, pota, pot͡ʃo, kat͡ʃo/ and

[6] Where sources differ on which Gomera Spanish phonemes correspond to which Silbo Gomero whistlemes, we have relied on Rialland (2005), whose work is based on perceptual experiments.

so on (Trujillo 1978: 148).

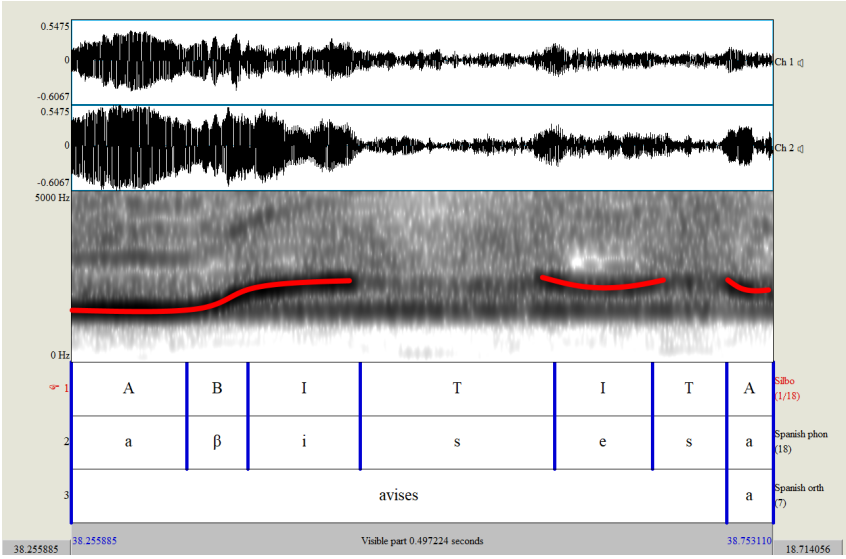
Even with the heavily reduced inventory of Silbo Gomero, there is evidence that it shares a phonological grammar with Gomera Spanish. This is evident in processes applying in both speech and whistling, such as s-deletion and m-to-n, which apply in codas in Gomera Spanish. Below in (6) is a whistled and spoken example of s-deletion from Classe (1957), which shows up as T-deletion in Silbo Gomero. Classe (1957: 65) is explicit that all phonological processes "are maintained in the whistled form of Gomera Spanish."

(6) Gomera Spanish S-deletion corresponds to Silbo Gomero T-deletion

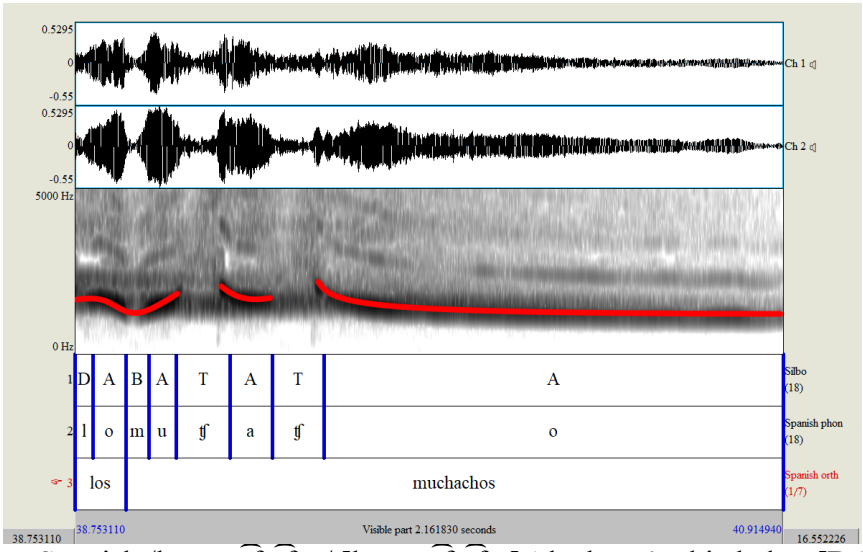
Underlying form	Gomera Spanish	Silbo Gomero	Translation
/as bisto/	[a βito], not [as bisto]	[A BITA], not [AT BITTA]	'you (sg.) have seen'

The same morpheme may show up with or without an [s] depending on the phrasal environment: if a word-final /s/ is followed by a pause, it is deleted, but if the /s/ is followed by a vowel-initial word, it surfaces as [s] (Lipski 1985: 128). Below are examples containing word-final /s/ in deleting and non-deleting positions from the same whistler and the same utterance (audio from Wikimedia Commons):

(7) Word-final /s/ in Silbo Gomero



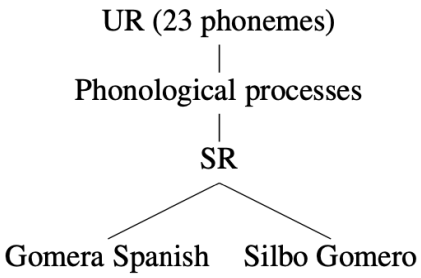
a. Gomera Spanish /abises a/ [aβises a] 'tell (to)' whistled as [ABITIT A], with [T]



b. Gomera Spanish /lo̲s mut͡ʃat͡ʃo̲s/ [lo̲ mut͡ʃat͡ʃo̲] 'the boys' whistled as [DA̲ BATATA̲], without [T]

We model this situation as follows:

(8)



That is, the entire phonological grammar is the same for both varieties, as is suggested by the fact that all phonological processes are shared. The identical surface forms are then transduced in one of two different ways, to produce the articulatory instructions that result in speech and whistling respectively. This situation of having an identical phonology with non-identical phonetic outputs is not predicted by theories where the phonology-phonetics mapping is invariant. One way around this might be to posit a separate transducer for spoken and whistled speech, but in an innatist model this would entail that every speaker has a whistling mode, even if they do not know a whistled language. Adding to the difficulty of sustaining this position, there are no attested languages that are *only* whistled, so whistling is not a full-fledged separate modality, unlike the independent visual modality of signed languages (anticipating the next subsection). Why should a separate, universal transducer and set of phonological primitives be posited for a modality that is always parasitic on the grammar of a spoken

language?

A further problem with the innatist view can be illustrated with spoken and whistled featural mismatches, which are cases where both speech and whistling make use of the same phonetic contrast for different phonological purposes. Both Gomera Spanish and Silbo Gomero have phonemic contrasts based on continuancy. In Gomera Spanish, this is the familiar [continuant] feature, distinguishing the [+continuant] vowels, sonorants, and fricatives from the [-continuant] stops. In Silbo Gomero, the contrastive pairs /B/-/P/ and /D/-/T/ are only distinguished by continuancy. Both pairs share the same pitch movements, but while /B/ and /D/ involve continuous whistling, /P/ and /T/ contain a non-continuous period of silence in the middle, similar to the closure portion of a spoken stop. (Trujillo 1978, Rialland 2005). Despite this similarity in acoustic cues, whistled continuancy is unrelated to the [continuant] feature in speech, and instead maps onto spoken [voice], as noted by Rialland (2005: 249). The whistled non-continuants represent all and only voiceless phonemes, whether continuous (/f, s, t͡ʃ, χ/) or not (/p, t, t͡ʃ, k/). Gomera Spanish and Silbo Gomero are phonologically the same language, and both use binary acoustic contrasts based on continuancy. If there were an invariant mapping between phonetic properties and phonological contrasts, both the whistled and spoken continuants would encode the same contrasts. Instead, there is a mismatch, where the same phonetic cue corresponds to different phonological categories. This suggests that there is no (I-)language-invariant correspondence between phonetic properties and phonological contrasts.

2.4.2 Sign language

The wealth of research on sign language phonological feature systems provides an interesting contrast to spoken language in several ways. Most pressingly for the current work, the hypothesis that there is an innate feature set with acoustic and/or articulatory correlates for spoken languages raises the question of how this feature set might pertain to visual and tactile signed languages. Mielke (2008:16) raises three possibilities and provides several arguments, which we will review here in turn, for rejecting the first two:

> (1) relax the requirement that features are defined in phonetic terms and interpret each innate feature as having both spoken language and signed language phonetic correlates, (2) posit additional innate features which apply to signed language, and claim that humans are hardwired with two sets of innate features for two different modalities, or (3) consider that features and their phonetic correlates are learned during acquisition, according to the modality of the language being acquired.

The first possibility is compatible with an innatist view in which phonetics is language-specific such that a single feature may be transduced into different articulatory instructions, but not one in which transduction is invariant. If this were the case, we should see a strong correspondence between signed and spoken language features: they should be similar in number, and similar in feature geometric organization. Corina & Sagey (1989) reached the conclusion that this hypothesis is not borne out by studies of sign languages. For example, Stokoe et al. (1976) propose 19 handshapes, 12 places of

articulation, and 24 types of movement for ASL. Some models involve nearly 300 distinctive features (Liddell & Johnson 1989) or employ features with more than two values (Brentari 1990). Concurring with Corina & Sagey, Sandler (2014:193) states that these feature sets "bear no relationship to those of spoken language." Moreover, feature geometric models (i.e., feature hierarchies) for sign (e.g. Sandler 1989, Corina & Sagey 1989) have little resemblance to those proposed for spoken languages. At the syllable level, too, there appear to be significant differences between sign and speech: there is "nearly complete consensus across models of sign language phonology /.../ that movements are the nuclei of the syllable" (Brentari 2011: 695), but movement is generally predictable from the start and end locations of the syllable. Some feature models of sign languages therefore reject movement as a primitive, instead deriving it from the start and endpoints (Liddell & Johnson 1989, van der Hulst 1993). This would be equivalent to a spoken language allowing only a single contrastive segment in the nucleus, clearly far from the norm in spoken languages. Due to the striking lack of similarity between spoken and signed languages, the possibility that they share a single feature system and differ only in phonetic implementation can be rejected.

Mielke (2008) suggests that the second possibility seems unlikely from an evolutionary perspective, since deafness is relatively rare and sporadic, and all known sign languages are young; deaf communities were likely only established in the past few hundred years, and no hearing communities appear to use sign as their primary modality (Sandler 2012, Sandler 2014). The same would hold for tactile signs used by the deaf-blind, all the more so since there do not appear to be any languages that are exclusively conveyed in the tactile modality. As Sandler (2014:184) puts it, "spoken language was evolution's choice." Some researchers have suggested that language may find its evolutionary origins in gesture (e.g. Stokoe 2002, Tomasello 2008, Arbib 2012; see Kendon 2017 for a critical overview). However, none of these proposals go so far as to assume that these gestural proto-linguistic systems had duality of patterning, a necessary precondition for phonological features (Sandler 2014). To our knowledge, there are no supporters on record for the idea that sign language and spoken language have innate but distinct feature systems.

This leaves us with the third possibility, that phonological features of visual signed languages are acquired. It would seem to follow by the same logic that the phonological features of tactile signs and whistled languages would also be acquired. Put simply, if we can acquire features for these modalities, why not for speech? This seems to us to be the most likely origin of phonological features for sign, and as such provides support the position that features are learned for speech as well.

3. Learning features and 'substance-freedom'

In substanceful theories of phonological computation, innate rules or constraints may refer to phonological features (as in most versions of Optimality Theory; see e.g. Tesar & Smolensky 2000). Under such a conception of the language faculty, features must be innate. If phonological computation is substance-free, though, then there exists a logical possibility that features are acquired. On this view, not only are there no innate features,

the features learned by each child don't necessarily correspond to phonetic divisions of a language's phoneme inventory at all. Archangeli and Pulleyblank (2015: 2) describe these as "categories": "these categories correspond roughly to the familiar 'distinctive features,' though there is no *a priori* set of features to map the sounds to, and in fact, a behavioral category is not necessarily an acoustic or articulatory category, and vice versa." As noted in (1), theories in which featural content is learned differ in whether substance is involved in acquisition of features (phonetic feature-learning) or not (phonological feature-learning). We review the evidence in favor of each of these views in the subsections to follow.

3.1 Approaches to learning featural content

3.1.1 Arguments that featural content is learned from phonology

The phonological feature-learning view is perhaps best exemplified by Mielke's (2004, 2008) emergent feature theory, Morén (2007), Nazarov (2014), and Odden (this volume). The main argument for a phonological feature-learning view is an empirical one, outlined by Mielke (2004 et seq.). Mielke argues that the phonological systems of the world's languages need to refer to a phonetically 'unnatural' class as the target or environment of a phonological process more often than phonologists have previously thought. Out of 6,077 phonological classes in Mielke's database from 628 language varieties, 1,498 of them are unnatural according to all three of the most historically influential feature theories: those proposed by *Preliminaries of Speech Analysis* (Jakobson et al. 1952) and *The Sound Pattern of English* (*SPE*, Chomsky and Halle 1968) and the more recent Unified Feature Theory (Clements and Hume 1995).

For example, in River West Tarangan (Nivens 1992: 219, Mielke 2008: 121), /m/ undergoes place assimilation to a following /t̪ g s j/, but no other segments:

(9) Place assimilation of /m/ in River West Tarangan
   a. /bimtem/ → [bi**n**tém] 'DUP small'
   b. /jɛrgimgum/ → [jɛrgi**ŋ**gum] 'DUP NF rub'
   c. /simsimə/ → [si**n**simə] 'ant (sp)'
   d. /ɸaɸamjɛmnə/ → [ɸaɸa**n**jemnə] 'overcast 3s'
   e. /jɛrkimkam/ → [jɛrkimkam] 'DUP NF dislike'
   f. /dimdumdi/ → [dimdumdi] 'DUP six PL'

These processes apply across morpheme boundaries, so must be productive in some part of the grammar. Given the consonant inventory of River West Tarangan, there is no combination of features under any of Jakobson et al. (1952), *SPE*, or Unified Feature Theory that could encode the set /t̪ g s j/ as a phonological natural class:

(10) River West Tarangan consonant inventory (Mielke 2008: 122)

	t̪	k
b	d	**g**
ɸ	**s**	
m	n	ŋ
	r	
	l	
	j	

There are a few possible responses to this set of data. One is to claim that there are no featurally unnatural phonologically active classes, and find other analyses of the phonemic inventory of the language to describe every unnatural-looking class featurally. The resulting feature may not have any identifiable acoustic or articulatory correlate. This would be a phonological feature-learning approach. We refer the reader to Odden (this volume) for a concrete proposal concerning how this learning would proceed.

A second response, which could be made in an attempt to salvage the innatist position, would be to argue that River West Tarangan has four separate rules, causing place assimilation of /m/ before /t̪/, and separately before /g/, and separately before /s/, and separately before /j/. This potential analysis has some merit, given that assimilation is obligatory before /t̪/ but optional before /g s j/ (Mielke 2008: 121); something special has to be said about the status of /t̪/, though notice that the class is featurally unnatural with or without this segment. Reiss (2017) argues that children can only formulate rules that refer to the featural intersection of the segments that participate in the observed alternation. The four-rule analysis is thus unavailable; it is not a learning strategy provided by UG. However, in such a scenario the child must have some way to retreat from overgeneralizing and revert to multiple rules. Perhaps, upon realizing that feature intersection would predict the participation of other [+consonantal] segments known not to participate, the child avoids doing so.

In short, there is a decision to be made in light of languages like River West Tarangan: allow rules to proliferate, or allow features to be created. In light of the arguments presented in the previous section, we suggest the latter. This is also relevant to phonetic feature-learning approaches discussed below, which will need to permit the phonology to "override" the phonetics in these cases, as Dresher (2018) puts it. Indeed, Sapir (1925: 47-48) already noted that "it is most important to emphasize the fact, strange but indubitable, that a pattern alignment does not need to correspond exactly to the more obvious phonetic one."[7]

[7] However, some cases claimed to illustrate phonetically unnatural patterns *are* dubious and underscore the need to be cautious about the types of data that are brought to bear on this issue. See Hall (2010) and Uffmann (2018) for further discussion.

3.1.2 Differences among phonological feature-learning approaches

Few phonological feature-learning approaches have been explicated in sufficient detail to test their predictions, but one way of distinguishing between approaches such as Dresher's (2014, 2018) from Odden's (this volume) lies in cases where non-contrastive features appear to be needed in the representation. Dresher (2014: 166) states that "it is the learners' task to arrive at a set of features that account for the contrasts in the lexical inventory (the phonemes) of their language." As we understand Odden's approach, active features are first learned on the basis of phonological alternations, and non-active but contrastive features are filled in later. This may result in differing featural representations of segments and different geometric organizations of the features.

Pulleyblank (2003), Hall (2007), and Nevins (2015) discuss a number of cases that pose a concern for contrast-based feature assignment, as formalized e.g. in Hall's (2007: 20) Contrastivist Hypothesis, which states that "[t]he phonological component of a language L operates only on those features which are necessary to distinguish the phonemes of L from one another." Rather than abandoning the hypothesis entirely, Hall proposes the 'prophylactic' assignment of non-contrastive features when necessary. We will illustrate the general nature of the argument with Czech, following Hall (2007: 40ff). Czech has final devoicing, as in (11):

(11) Final devoicing in Czech (Hall 2007)

a.	muž	[muʃ]	'man' (nom.sg)
b.	mužem	[muʒem]	'man' (inst.sg)
c.	myš	[miʃ]	'mouse' (nom.sg)
d.	myši	[miʃi]	'mouse' (inst.sg)

Czech also has regressive voicing assimilation in consonant clusters when the second member is an obstruent, both within words and across word boundaries. This can be demonstrated as in (12) with the prepositions /s/ 'with' and /z/ 'from'.

(12) Regressive voicing assimilation before obstruents, not sonorants

a.	s lesem	[slesem]	'with a forest'
b.	z lesa	[zlesa]	'from a forest'
c.	s mužem	[smuʒem]	'with a man'
d.	z muže	[zmuʒe]	'from a man'
e.	s domem	[zdomem]	'with a house'
f.	z domu	[zdomu]	'from a house'
g.	s polem	[spolem]	'with a field'
h.	z pole	[spole]	'from a field'

Two segments, /v/ and /r̩/, undergo final devoicing and regressive assimilation in these contexts, as is to be expected.

(13) Final devoicing of /v/ and /r̝/

a.	zpěv	[spjef]	‘song’ (nom.sg)
b.	zpěvem	[spjevem]	‘song’ (inst.sg)
c.	barev	[baref]	‘colors’ (gen.pl)
d.	barva	[barva]	‘color’ (nom.sg)
e.	lékař	[le:kar̝̊]	‘doctor’ (nom.sg)
f.	lékařem	[le:kar̝̊em]	‘doctor’ (inst.sg)

(14) Regressive assimilation of /v/ and /r̝/

a.	v lese	[vlese]	‘in a forest’
b.	v pole	[fpole]	‘in a field’
c.	v chybě	[fxibje]	‘in a mistake’
d.	nářek	[na:r̝̊ek]	‘lamentation’ (nom.sg)
e.	nářky	[na:r̝̊ki]	‘lamentations’ (nom.pl.)

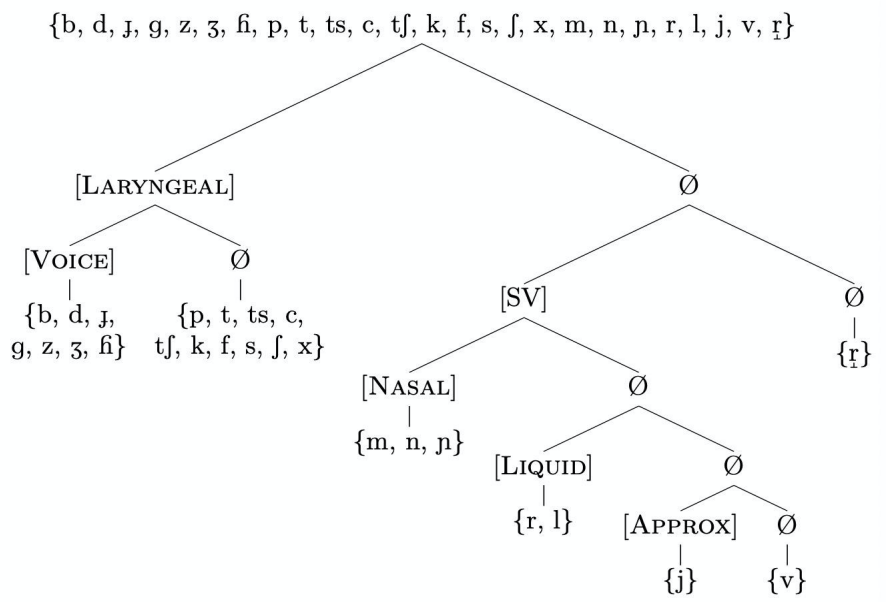
Interestingly, neither /v/ nor /r̝/ triggers voicing assimilation when it is the second segment in a consonant cluster, as demonstrated below. See Hall (2007) concerning dialectal variation, but for the present purposes, it will suffice to discuss dialects in which /v/ remains voiced but /r̝/ is voiceless if it is adjacent to a voiceless obstruent, as we have already seen in the previous examples.

(15) Non-triggering of assimilatory voicing by /v/ and /r̝/

a.	s vránou	[svra:nou̯]	‘with a crow’
b.	květ	[kvjet]	‘flower’
c.	středa	[str̝̊eda]	‘Wednesday’
d.	při	[pr̝̊i]	‘near’

The problem is as follows: Hall adopts the Successive Division Algorithm (SDA), which begins by assuming that an early language learner “can perceive and distinguish speech sounds, but has not yet identified any contrasts” in the phonological system (Dresher 2014: 172). Thus, the learner begins with a single undifferentiated phoneme and then, upon finding a phonological contrast, “[assigns] contrastive features by successively dividing the inventory until every phoneme is distinguished” (Dresher 2014: 166; see Dresher 1998, 2009 for a more explicit algorithm). Thus, all and only the phonologically contrastive features are assigned, creating a language-specific, non-innate hierarchical organization. Hall (2007) argues that, in Czech, the SDA will produce a featural hierarchy like the one below (place features omitted for brevity):

(16) Consonant inventory of Czech (reproduced from Samuels 2009: 58, following Hall 2007: 82)



According to the feature geometry in (16), we can describe voicing assimilation as [Laryngeal] spreading leftward. The sonorants, including the "lapsed sonorant" or "sonorant obstruent" /v/ (Hall 2007: 48, 54), are not specified with [Laryngeal] so they do not trigger this process. However, the [SV] (Sonorant or Spontaneous Voicing) feature of /v/ can be 'overridden' by [Laryngeal] and its dependent [Voice] spreading. Note also that [Laryngeal] is a phonological feature with no phonetic correlate: "the feature [Laryngeal] is associated not with any articulatory, auditory, or acoustic attribute of the sounds to which it is assigned," but instead by its ability to trigger voicing assimilation (Hall 2007: 56).

The problem is that /r̝/ does not bear either [Laryngeal] or [SV].[8] When /r̝/ is in a position to be devoiced, such that [Laryngeal] spreads onto it, the result should be that it turns into an obstruent specified only for [Laryngeal] and nothing else. This would not be /r̝/, because /r̝/ is not a phoneme. This analysis therefore incorrectly predicts that a devoiced /r̝̊/ should surface as some other phoneme, say /t/. Hall's solution is to specify /r̝/ for some non-contrastive feature--any one will do, because it need not be visible to the phonology. It does not trigger, block, or undergo any phonological processes. It simply prevents spreading of [Laryngeal] to /r̝/ from giving the wrong result because it keeps /r̝/ distinct from /t/ (or whatever the maximally underspecified obstruent happens to be) when it's devoiced. Dresher admits only contrastive features in the phonology, though;

[8] A maximally underspecified segment like this one necessarily occurs in every inventory produced by the SDA, so this problem is not an idiosyncrasy of Czech; see Hall (2007, section 1.2.7).

beyond that, only 'post-phonological' (phonetic) enhancement is permitted. The phonological portion of the theory would need to be augmented in order to account for Czech and cases like it. On the other hand, the analysis of voicing assimilation crucially depends on [Laryngeal], a feature that can only be identified by its phonological behavior, not the phonetic properties of segments that bear it.

In contrast to the SDA-based feature assignment described above, in Odden's approach, we would begin by assigning a feature--call it A--to the segments that undergo final devoicing. We would then assign another feature, B, to the segments that undergo regressive assimilation, and C to segments that trigger it. Since /v/ and /r̥/ undergo final devoicing and regressive assimilation but do not trigger regressive assimilation, they are both specified {A, B} but not {C}. As we have seen, /v/ stays voiced whereas /r̥/ undergoes progressive assimilation (i.e., devoicing), so /r̥/ gets another feature, D, to indicate this. This process of assigning features appears to provide a straightforward description of the facts, though it does not make the predictions about the range of possible phonologies entailed by adopting the Contrastivist Hypothesis.

The phonological feature-learning approach of Nazarov (2014: 24) differs considerably from the others discussed here, in that "the [acquisition] of phonological features is not motivated by an explicit requirement to have phonological features per se. Instead, features are induced by the learner because they help state constraints in a way that generalizes over more forms." Constraints and features are induced in an iterative process, and constraints can refer to either features or directly to pre-existing "atomic segment units," which he maintains "are not shorthand for a feature bundle" (Nazarov 2014: 23). In Nazarov's model, the drive to maximize generalization will cause learners to state constraints over features only when multiple segments are affected--a position similar to that of Reiss (2017a). Thus, claims Nazarov (2014: 32), the learner will not generalize constraints that refer to a single segment: for example, a hypothetical constraint *m# banning [m] word-finally will not apply to novel [labial, nasal] segments. It is not clear, though, that this is a desirable prediction—or that it actually differs from the predictions made by Reiss (2017a), Odden (this volume), or indeed any other model presented here. Say for the sake of argument that the repair for word-final [m] is deletion. A Reissian child would trivially take the intersection of all the segments that undergo this process and arrive at the feature matrix for [m] (fully specified, since Reiss subscribes to an 'archiphonemic' view of underspecification in which only segments that alternate are underspecified; see Reiss 2008). The featural representation of a novel [labial, nasal] segment would differ from that of [m] and therefore not satisfy the conditions for deletion. An Oddenian child would assign some arbitrary feature to [m], and only [m], as it alone undergoes word-final deletion. Again, this would not generalize because a novel segment would not be assigned this feature without positive evidence. Of course, whether an *actual* child would fail to generalize to a novel [labial, nasal] segment should be tested experimentally. We discuss predictions regarding generalization further in 3.3.

3.2 The phonetic feature-learning view

Under the phonetic feature-learning view (e.g. Niyogi 2004, Lin 2005, Lin & Mielke 2008, Beguš 2020, Cui 2020), features are acquired from the input and are not innate; but unlike under the phonological feature-learning view, this acquisition process is guided by phonetic substance. This approach takes seriously Hyman's (1975: 25) pronouncement that "When a phonetic property can be extracted, a generalization is revealed. When no phonetic property can be extracted, these segments should not be able to occur as a class in linguistics." A number of different models for mapping acoustic data to phones have been proposed (e.g. de Boer 2001, Niyogi 2004, Oudeyer 2006, Coen 2006 and many others), though this literature tends to be agnostic regarding whether the categories to be learned are phonemic or phonetic (see Dillon et al. 2013 on this point) and few commit to any particular model of phonology. Nevertheless, it's important to point out that phonology under the phonetic emergentist view must still be substance-free, even though the acquisition process is guided by substance: learned phonological categories are incompatible with theories in which there are innate rules or constraints that refer to phonetics, e.g. phonetically grounded theories.

Lin (2005: 74ff) proposes a model using unsupervised, iterative learning methods, incorporating a sub-model for identifying segments from a waveform followed by a sub-model for partitioning these segments to acquire features. Interestingly, Lin (2005: 98ff) also suggests that it is feasible to learn features inductively directly from the word-level acoustic waveform, and that the categories learned in this manner are very similar to those learned from the same training data augmented with segmental boundary information.

Another recent example of a phonetic learning-based approach is that of Cui (2020), who proposes a model of category acquisition in which the learner creates acoustic features one by one when needed to represent a lexical contrast. The child starts off with no *phonological* contrasts, representing individual acoustic tokens that all share the same phonological representation. When the learner finds evidence that two of the words in their lexicon have distinct referents, they pick the most salient cue distinguishing the two tokens, and turn that cue into a phonological feature. The feature defines a plane through acoustic space, and after the first division event, the child now has two categories. New tokens are sorted into those two categories according to which side of the plane they fall on.

Whenever the child discovers a new lexical distinction not representable by their current feature system, they again create a new feature defined by the most salient acoustic cue distinguishing the two tokens, and then use that cue to parse all future inputs. The phonetic category acquisition process stops when the child has enough features to represent every lexical contrast in the language. In this model, features are strictly acoustically defined: all featural distinctions correspond directly to planes in acoustic space. Features can be acquired in a language-specific (and indeed speaker-specific) order (similarly to the more phonologically driven proposal in Dresher 2018), with the most salient acoustic cues that distinguish the most frequent words in the language becoming categories the earliest.

Beguš (2020) uses raw acoustic data to train a Generative Adversarial Network on an aspect of the allophonic distribution of English voiceless stops (aspirated word-initially if followed by a stressed vowel; unaspirated after [s]). The network was able to learn and generalize this generalization as well as to identify phonetic variables that correspond to the presence of [s]. This approach is similar to that of Dillon et al. (2013) in that it learns aspects of phonetics and phonology at the same time. However, more work is needed to establish that this approach is feasible for other phonological alternations; opaque patterns and phonetically unnatural patterns would be particularly interesting cases.

3.3 Phonetics, phonology, or both?

The feature-learning approaches form a spectrum from those based entirely on phonology (e.g., Odden, Morén, Nazarov) to those that rely primarily on phonology but explicitly supplement it with phonetics (e.g., Archangeli & Pulleyblank, Dresher) through to those based exclusively on phonetics (e.g., Lin, Cui). As the example of River West Tarangan in 3.2 shows, we suspect that a purely phonetic approach will fail to account for many of the cases documented by Mielke (2008). It is also interesting in this regard that all the 'phonetic' feature-learning approaches allow phonology to overrule the phonetics.[9] There is a division of labor and a logical sequence of events implied by both the phonetic and phonological feature-learning models: first the child identifies segments pre-phonologically on a phonetic basis, and then identifies the features that play a role in the phonological computation (i.e., are active). If this is correct, then the role of phonological features is purely computational, not to define contrasts (cf. Reiss 2017b).

This is why we argue phonetic and phonological feature-learning approaches are stronger together. Lin (2005) and Dillon et al. (2013) point out that most approaches to phonological acquisition assume segmental (phonemic) encoding of speech as their input. This is evident in the works reviewed in 3.1. Features, on this view, can be acquired by partitioning the segments based on their phonological activity, again as reviewed in the previous subsection. However, the starting assumption raises serious questions: how do children come to be able to parse the speech stream into segments? Moreover, what *are* these pre-existing segments, if not bundles of features? This is a particularly salient question for Nazarov, whose theory can refer directly to "atomic segment units." We should not be fooled by the use of IPA into thinking that "atomic segment units" are the same type of thing as feature bundles that we also happen to label with IPA symbols as shorthand. It is very difficult to maintain the position that a child at this early stage of learning will be able to assign their percepts to unambiguous phonological categories, as we know (recall 2.3) that these categories vary from language to language and can overlap considerably.[10] Perhaps the phonological feature-learning approaches are still on

[9] Krekoski (2013, 2017) and Dresher (2014) provide interesting evidence from Chinese tonal systems that phonological systems can remain stable despite phonetic changes, but only up to a certain point: when the phonetics no longer aligns with the phonological activity (in these cases, tone sandhi), the system is liable to undergo reanalysis. Such restructuring may weed out phonetically unnatural rules over time.

[10] This is a feature-learning version of the concern raised by Dresher (2015: 174) regarding Hale & Reiss's (2008) innatist proposal, namely that features are inherently ambiguous (Dresher asks: "How low qualifies

the right track, but represent a stage after which the child has already used a phonetic strategy to arrive at a set of phones. Alternatively, Dillon et al. (2013)[11] and Beguš (2020) propose models that simultaneously learn aspects of phonetics (phones for Dillon et al. and features for Beguš) and the phonological processes that produce allophonic variation. Although this family of approaches is interesting, they have been demonstrated with only a couple of allophonic rules thus far.

It also remains unclear whether phonological feature-learning approaches can be modified to make correct predictions regarding the degree to which phonological rules are generalized by the learner. In 3.1.2 we discussed the special case of phonological rules that apply to only one segment, but it is worth considering this issue more broadly. Experiments with both infants and adults have demonstrated that learners generalize phonological rules to new contexts that would be featurally natural in an innatist approach (e.g. Cristià & Seidl 2008, Finley & Badecker 2009). For example, Cristià & Seidl (2008) showed that infants will generalize a pattern involving nasals and stops to other members of the [-continuant] class. Speakers will also generalize a rule to a non-native/novel phoneme, as in the famous example of English speakers producing the plural of *Bach* /bɑx/ with [s] due to the voicelessness of /x/ (Halle 1978). Here a contrastivist approach like that of Dresher or Hall correctly predicts generalization (see Dresher 2015 on loanword adaptation, which poses a very similar issue). However, for reasons discussed in 3.1.2, it seems to us that Odden (this volume) will not capture such generalizations.

One crucial component that appears to be missing from several of the phonetic and phonological feature-learning approaches we have discussed thus far is the ability to generate a phonemic system in addition to generating featural representations for all the surface phones. Odden (this volume) explicitly denies that 'taxonomic phonemes' need to be accounted for, but others remain silent. To the extent that generating a systematic phonemic inventory is desirable, phonological feature-learning models will need to be augmented by a second step to do so. Again here the approach taken by Dillon et al. (2013) differs from the rest, as it is able to achieve all of this in one step. Perhaps this proposal could be coupled with a phonological or mixed phonetic-phonological feature-learning approach to learning phonemes, allophones, and features—but currently, no model does it all.

4. Conclusion and future directions

The spectrum of proposals regarding the status of phonological features within SFP is broad, and it is even broader when one considers models of phonological acquisition that are not explicitly substance-free but describe unsupervised learning methods for discretizing and categorizing acoustic input into phones. Despite this, we still find considerable common ground. With the exception of Nazarov, we all seem to agree--or at

as [low]?") and it seems unwise to assume that a learner could arrive at an unambiguous featural representation for the speech sounds they hear. See Cowper & Hall (2014) for further helpful discussion.

[11] Dillon et al. remain agnostic about whether features are innate, or if not, how they would be acquired (see discussion in their section 5.2).

least, no other works of which we are aware explicitly *dis*agree--that the notion "feature" is innate; that there is some innate syntax of features, albeit general enough to encompass both sign and speech; and that certain properties of our auditory processing system that are shared with other species appear to affect our phonological abilities. Among those who advocate feature-learning approaches, we also seem to agree that the procedure by which children go about acquiring features is innate. Although we can only speak for ourselves, this would seem to situate us all in the nativist tradition. We might disagree about the initial premises of Hale & Reiss's card grammar analogy, but we don't deny the basic argument: in their words, "ya gotta start with something!"

We have started--but much remains to be explored within the emergentist family of approaches. Future research will need to consider both theory and data carefully. Throughout Section 3 we have suggested a number of ways in which specific proposals are in need of refinement or further explication. For example, none of the phonetic theories address what the learner should do when confronted with a phonetically unnatural but phonologically active class, and the mixed theories are not explicit about the roles of each of these types of evidence. None of the theories discussed here except Dillon et al. (2013) generate a phonemic inventory, but although they do not presuppose features, their theory doesn't explain how features are acquired, nor do many of the other models that learn phonetic categories from acoustic waveforms.

Though we have presented them as alternatives here, it seems that the phonological and phonetic feature-learning approaches could and should be synthesized, particularly in light of the fact that the phonological approaches seem to begin *in medias res* with a learner who can identify phones from the acoustic stream, whereas the phonetic approaches tackle this earlier stage of learning. Moreover, and very unfortunately given the arguments presented in 2.4, none of these proposals address learning features in modalities other than speech, something our theory of phonological acquisition must account for. We believe that these directions for future research will lead to more explicit, testable models, which will in turn strengthen SFP's position as a comprehensive theory of phonological competence from infancy to adulthood.

References

Andersson, Samuel. In press. I'm sorry but time travel isn't real: against counterfeeding from the past. *Penn Working Papers in Linguistics* 26(1).

Anttila, Arto. 2007. Variation & optionality. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, 519–536. Cambridge University Press.

Arbib, Michael A. 2012. *How the brain got language: the mirror neuron hypothesis*. Oxford: Oxford University Press.

Archangeli, Diana & Douglas Pulleyblank. 2015. Phonology without universal grammar. *Frontiers in Psychology* 6. 1229.

Beguš, Gašper. 2020. Modeling unsupervised phonetic & phonological learning in Generative Adversarial Phonology. *Proceedings of the Society for Computation in Linguistics* 3. Article 15.

Bender, B. 1968. Marshallese phonology. *Oceanic Linguistics* 7. 16–35.

Bent, T., A.R. Bradlow & B.A. Wright. 2006. The influence of linguistic experience on pitch perception in speech & non-speech sounds. *Journal of Experimental Psychology: Human Perception & Performance* 32(1). 97–103.

Berent, Iris, Evan Balaban, Tracy Lennertz & Vered Vaknin-Nusbaum. 2010. Phonological universals constrain the processing of nonspeech stimuli. *Journal of Experimental Psychology: General* 139(3). 418–435.

Blaho, Sylvia. 2008. *The syntax of phonology: a radically substance-free approach.* University of Tromsø Ph.D. dissertation.

de Boer, Bart. 2001. *The origins of vowel systems*. Oxford: Oxford University Press.

Brentari, Diane. 1990. Licensing in ASL handshape. In C. Lucas (ed.), *Sign language research: theoretical issues*, 57–68. Washington, DC: Gallaudet Uni- versity Press.

Brentari, Diane. 2011. Sign language phonology. In John A. Goldsmith, Jason Riggle & Alan C.L. Yu (eds.), *Handbook of phonological theory*, 691–721. Oxford: Blackwell.

Brown, Charles H. & Joan M. Sinnott. 2006. Cross-species comparisons of vocal perception. In Steven Greenberg & William A. Ainsworth (eds.), *Listening to speech: an auditory perspective*, 183–201. Mahwah, NJ: Lawrence Erlbaum Associates.

Carreiras, Manuel, Jorge Lopez, Francisco Rivero & David P. Corina. 2005. Linguistic perception: neural processing of a whistled language. *Nature* 433(7021). 31–32.

Chabot, Alex & Tobias Scheer. 2019. What is it that is substance-free: computation and/or melodic primes. http://facultyoflanguage.blogspot.com/2019/03/a-new-player-has-entered-game.html. Accessed July 2, 2020.

Choi, J. 1992. Phonetic underspecification & target interpolation: an acoustic study of Marshallese vocalic allophony. *UCLA Working Papers in Phonetics* 82.

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

Classe, A. 1957. Phonetics of the Silbo Gomero. *Archivum Linguisticum* 9. 44–61.
Clements, George N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 245–306. Oxford: Blackwell.

Coen, Michael Harlan. 2006. *Multimodal dynamics: self-supervised learning in perceptual & motor systems*. Cambridge, MA: MIT dissertation.

Corina, David P. & Elizabeth Sagey. 1989. Are phonological hierarchies universal? Evidence from American Sign Language. In K. de Jong & Yongkyoon No (eds.), *Proceedings of the 6th Eastern States Conference on Linguistics*, 73–83.
Cowper, Elizabeth & Daniel Currie Hall. 2014. Reductiō ad discrimen: where features come from. *Nordlyd* 41(2). 145–164.
Cristià, Alejandrina & Amanda Seidl. 2008. Is infants' learning of sound patterns constrained by phonological features? *Language Learning & Development* 4(3). 203–227.
Cui, Aletheia. 2020. *The emergence of phonological categories*. University of Pennsylvania dissertation.
Dillon, Brian, Ewan Dunbar & William J. Idsardi. 2013. A single-stage approach to learning phonological categories: insights from Inuktitut. *Cognitive Science* 37(2). 344–377.
Dooling, Robert, Kazuo Okanoya & Susan Brown. 1989. Speech perception by budgerigars (*Melopsittacus undulatus*): the voiced–voiceless distinction. *Perception & Psychophysics* 46. 65–71.
Dresher, B. Elan. 1998. Child phonology, learnability, & phonological theory. In Tej Bhatia & W.C. Ritchie (eds.), *The handbook of language acquisition*, 299–346. New York: Academic Press.
Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.
Dresher, B. Elan. 2014. The arch not the stones: universal feature theory without universal features. *Nordlyd* 41(2). 165–181.
Dresher, B. Elan. 2015. The motivation for contrastive feature hierarchies in phonology. *Linguistic Variation* 15: 1-40.
Dresher, B. Elan. 2018. Contrastive feature hierarchies in synchronic and diachronic phonology. Phonological Studies (Journal of the Phonological Society of Japan) 21, 91–98.
Finley, Sara & William Badecker. 2009. Artificial language learning & feature- based generalization. *Journal of Memory & Language* 61(3). 423–437.
Hale, Mark. 2000. Marshallese phonology, the phonetics-phonology interface, & historical linguistics. *The Linguistic Review* 17. 241–257.
Hale, Mark, Madelyn Kissock & Charles Reiss. 2007. Microvariation, variation, & the features of universal grammar. *Lingua* 117(4). 645–665.
Hale, Mark & Charles Reiss. 2000. Substance abuse & dysfunctionalism: current trends in phonology. *Linguistic Inquiry* 31(1). 157–169.
Hale, Mark & Charles Reiss. 2003. The subset principle in phonology: why the tabula can't be rasa. *Journal of Linguistics* 39. 219–244.
Hale, Mark & Charles Reiss. 2008. *The phonological enterprise*. Oxford: Oxford University Press.
Hall, Daniel Currie. 2007. The role and representation of contrast in phonological theory. University of Toronto PhD dissertation.
Hall, Daniel Currie. 2010. Probing the unnatural. *Linguistics in the Netherlands* 73–85.
Halle, Morris. 1978. *Formal versus functional considerations in phonology*.

Bloomington: Indiana University Linguistics Club.
Hay, Jessica. 2005. *How auditory discontinuities & linguistic experience affect the perception of speech & non-speech in English- & Spanish-speaking listeners*. Austin, TX: University of Texas dissertation.
Holt, Lori L., Andrew J. Lotto & Randy L. Diehl. 2004. Auditory discontinuities interact with categorization: implications for speech perception. *Journal of the Acoustical Society of America* 116. 1763–1773.
van der Hulst, Harry. 1993. Units in the analysis of signs. *Phonology* 10. 209–241. Hyman, Larry. 1975. *Phonology: theory & analysis*. New York: Holt, Rinehart, & Winston.
Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to speech analysis: the distinctive features & their correlates*. Cambridge, MA: MIT Press.
Jusczyk, Peter W., D.B. Pisoni, A. Walley & J. Murray. 1980. Discrimination of relative onset time of two-component tones by infants. *Journal of the Acoustical Society of America* 61(1). 262–270.
Keating, Patricia. 1984. Phonetic & phonological representation of stop consonant voicing. *Language* 60(2). 286–319.
Kendon, Adam. 2017. Reflections on the "gesture-first" hypothesis of language origins. *Psychonomic Bulletin & Review* 24. 163–170.
Kluender, Keith R., Randy L. Diehl & Peter R. Killeen. 1987. Japanese quail can learn phonetic categories. *Science* 237. 1195–1197.
Kohler, Klaus. 1999. German. In *Handbook of the International Phonetic As- sociation: a guide to the use of the International Phonetic Alphabet*, 86–89. Cambridge: Cambridge University Press.
Krekoski, Ross. 2013. On tone and the nature of features. Ms., University of Toronto.
Krekoski, Ross. 2017. Contrast and complexity in Chinese tonal systems. University of Toronto PhD dissertation.
Kuhl, Patricia K. & James D. Miller. 1975. Speech perception by the chinchilla. *Science* 190. 69–72.
Kuhl, Patricia K. & Denise M. Padden. 1982. Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics* 32. 542–550.
Lajard, M. 1891 [1976]. Le langage siffle des Canaries. In Thomas Sebeok & Donna Jean Umiker-Sebeok (eds.), *Speech surrogates: drum & whistle systems*, vol. 2, The Hague & Paris: Mouton.
Liberman, Alvin M., F.S. Cooper, D.P. Shankweiler & Michael Studdert-Kennedy. 1967. Perception of the speech code. *Psychological Review* 74. 431–461.
Liberman, Alvin M. & I.G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21. 1–36.
Liddell, Scott & Robert Johnson. 1989. American Sign Language: the phonological base. *Sign Language Studies* 64. 197–277.
Lin, Ying. 2005. *Learning features & segments from waveforms: a statistical model of early phonological acquisition*: University of California Los Angeles dissertation.
Lin, Ying & Jeff Mielke. 2008. Discovering place & manner features: what can be learned

from acoustic & articulatory data? In J. Tauberer, Aviad Eilam & Laurel MacKenzie (eds.), *Penn Working Papers in Linguistics*, vol. 14 1, 241–254. University of Pennsylvania.

Lipski, John M. 1985. Reducción de s y n en el español caraio en norteamerica. *Revista de Filologia de la Universidad de La Laguna* 4. 125–133.

Mayer, Connor & Robert Daland. 2019. A method for projecting features from observed phonological classes. Ms., University of California Los Angeles.

Mesgarani, Nima, Stephen V. David, Jonathan B. Fritz & Shihab A. Shamma. 2008. Phoneme representation & classification in primary auditory cortex. *Jour- nal of the Acoustical Society of America* 123(2). 899–909.

Meyer, Julien. 2005. *Description typologique et intelligibilite´ des langues siffle´es, approche linguistique et bioacoustique*: Université Lumière Lyon 2 dissertation.

Mielke, Jeff. 2004. *The emergence of distinctive features*. Columbus, OH: Ohio State University PhD dissertation.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Mielke, Jeff. 2012. A phonetically-based metric of sound similarity. *Lingua* 122(2). 145–163.

Morén, Bruce. 2007. Phonological segment inventories & their phonetic variation: a substance-free approach. Paper presented at the GLOW XXX Segment Inventories Workshop, Tromsø.

Nazarov, Aleksei. 2014. A radically emergentist approach to phonological features: implications for grammars. *Nordlyd* 41(1): 21-58.

Nevins, Andrew. 2015. Triumphs and limits of the Contrastivity-Only Hypothesis. *Linguistic Variation* 15(1): 41-68.

Nivens, Richard. 1992. A lexical phonology of West Tarangan. In Donald A. Burquest & Wyn D. Laidig (eds.), *Phonological studies in four languages of Maluku*, 127–227. Summer Institute of Linguistics & University of Texas at Arlington.

Niyogi, Partha. 2004. Towards a computational model of human speech perception. In *Proceedings of the Conference on Sound to Sense*, 208–222. MIT Press.

Oudeyer, Pierre-Yves. 2006. *Self-organization in the evolution of speech*. Oxford: Oxford University Press.

Reiss, Charles. 2017a. Substance free phonology. In Stephen J. Hannahs & Anna R. K. Bosch (eds.), *Handbook of Phonological Theory*, 425-452. London & New York: Routledge.

Reiss, Charles. 2017b. Contrast is irrelevant in phonology. In Bridget D. Samuels (ed.), *Beyond Markedness in Formal Phonology*, 23-46. Amsterdam: John Benjamins.

Rialland, Annie. 2003. A new perspective on Silbo Gomero. In M.J. Sole, D. Recasens & J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 2131–2134. Barcelona: Causal Productions.

Rialland, Annie. 2005. Phonological & phonetic aspects of whistled languages. *Phonology* 22(2). 237–271.

Samuels, Bridget. 2009. *The structure of phonological theory*. Cambridge, MA: Harvard

University PhD dissertation.
Samuels, Bridget. 2011. *Phonological architecture: a biolinguistic perspective*. Oxford: Oxford University Press.
Samuels, Bridget. 2012. The emergence of phonological forms. In Anna Maria Di Sciullo (ed.), *Towards a biolinguistic understanding of grammar. essays on interfaces.*, 193–213. Amsterdam: John Benjamins.
Sandler, Wendy. 2012. The phonological organization of sign languages. *Language & Linguistics Compass* 6(3). 162–182.
Sandler, Wendy. 2014. The emergence of the phonetic & phonological features in sign language. *Nordlyd* 41(2). 183–212.
Sapir, Edward. 1925. Sound patterns in language. *Language* 1: 37-51.
Sebregts, Koen. 2015. *The sociophonetics & phonology of Dutch r*: Utrecht University dissertation.
Stokoe, William. 2002. *Language in hand: why sign came before speech*. Washington, DC: Gallaudet University Press.
Stokoe, William, D.C. Casterline & C.G. Croneberg. 1976. *Dictionary of American Sign Language on linguistic principles, second edition*. Silver Spring, MD: Linstok Press.
Streeter, Lynn A. 1976. Kikuyu labial & apical stop discrimination. *Journal of Phonetics* 4. 43–49.
Suga, Nobuo. 1969. Classification of inferior collicular neurons of bats in terms of responses to pure tones, FM sounds & noise bursts. *Journal of Physiology* 200. 555–574.
Suga, Nobuo. 1973. Feature extraction in the auditory system of bats. In A. Moeller (ed.), *Basic mechanisms in hearing*, 675–744. New York: Academic Press.
Suga, Nobuo. 2006. Basic acoustic patterns & neural mechanisms shared by hu- mans & animals for auditory perception. In Steven Greenberg & William A. Ainsworth (eds.), *Listening to speech: an auditory perspective*, 159–182. Mahwah, NJ: Lawrence Erlbaum.
Tesar, Bruce & Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
Tomasello, Michael. 2008. *The origins of human communication*. Cambridge, MA: MIT Press.
Trujillo, Ramón. 1978. *El Silbo Gomero: análisis lingüístico*. Santa Cruz de Tenerife: Andrés Bello.
Uffmann, Christian. 2018. A natural turn of Evenks. Paper presented at the 26th Manchester Phonology Meeting.
Vaux, Bert. 2008. Why the phonological component must be serial and rule-based. In B. Vaux & A. Nevins (eds.), *Rules, Constraints, and Phonological Phenomena*, 20-60. Oxford: Oxford University Press.
Vaux, Bert & Bridget Samuels. 2005. Laryngeal markedness & aspiration. *Phonology* 22(4). 395–436.
Vieregge, William H. & A.P.A. Broeders. 1993. Intra- and interspeaker variation of r in Dutch. *EUROSPEECH* 267–270.

Volenec, Veno & Charles Reiss. 2017. Cognitive phonetics: the transduction of distinctive features at the phonology-phonetics interface. *Biolinguistics* 11. 251– 294.
Volenec, Veno & Charles Reiss. 2020. Formal generative phonology. Manuscript available in *Radical: A Journal of Phonology*.
Waters, R. & W. Wilson. 1976. Speech perception by rhesus monkeys: the voicing distinction in synthesized labial & velar stop consonants. *Perception & Psychophysics* 19. 285–289.
Wikimedia Commons. 2006. Exemple de langage sifflé "El Silbo" pratiqué sur l'île de "La Gomera" (Canaries). https://commons.wikimedia.org/wiki/file:silbo.ogg.
Williams, L. 1974. *Speech perception & production as a function of exposure to a second language*. Cambridge, MA: Harvard University dissertation.