How Likely are Impossible Languages?

/t/ Epenthesis Reconsidered

Rebecca L. Morley

This paper provides re-analysis of 56 patterns of potential consonant epenthesis. The results call into question a hierarchy of context-free markedness constraints preferring /t/ to /k/ (Prince and Smolensky 1993/2004). In the sample, coronal and velar segments are equally likely; /t/, /g/, and /ʔ/ epenthesis languages occur exactly once each, /j/ epenthesis, twice. These results rely on establishing the best analysis of each pattern, and highlight the difficulty (and necessity) of separating linguistic data from theory. In the sample, conditioned glide epenthesis is relatively common, but default epenthesis, rare. A model linking synchronic patterns to their diachronic sources is proposed.

1. Introduction

Standard Optimality Theory posits a set of innate constraints that ban (violably) certain surface structures (Prince and Smolensky 1993/2004). Epenthesis is typically taken as a language-specific solution to satisfy a universal dispreference for syllables without onsets (e.g., input /ru.a/ realized as output [ru.ta]). That is, highly-ranked ONSET forces a faithfulness violation, and low-ranked DEP determines that the repair will involve insertion of a segment not present in the input. This formalism is blind to the origin of such a synchronic pattern and treats all such repairs equivalently. Under this theory, epenthesis is no different than deletion, metathesis, or coalescence. However, it has been noted previously that type of repair is, in fact, highly dependent on phonological context – the Too-Many-Solutions problem (e.g., Wilson 2001, Myers 2002).

At the same time, current generative theory may under-predict certain types of epenthesis patterns. A subset of the universal constraint set is typically instantiated as un-rerankable context-free markedness constraints. Of particular relevance to this paper are those that enforce an implicational hierarchy over place of articulation for inventories of plosives. If a language has /k/, then it also has /p/; if a language has /p/, then it also has /t/. This formalism also predicts /t/ as the preferred segment for unconditioned epenthesis, that is, as the default segment (cf de Lacy 2006[1]).

This paper is a re-examination of synchronic epenthesis via a typological study, an investigation of the parameters of phonological analysis, and a diachronic model of epenthesis-instantiating grammars. In a sample of 56 unique phonological patterns, all of which have been cited as epenthesis, a maximum of 22 languages is found to meet the

[1] de Lacy (2006) has an alternate formulation of the relevant constraints but arrives at the same conclusion in this case.

most basic requirements for true (that is 'default') epenthesis. That number is halved when more stringent evidential requirements are added. A relatively conservative analysis of the data suggests the following: that place of articulation is largely irrelevant to determining the quality of an epenthetic segment (with the notable exception of glides), and that 'default' epenthesis is, as a phonological phenomenon, uncommon. It will be argued that these facts are best captured by means by a framework that explicitly combines diachronic and synchronic mechanisms. The historically-based account allows for more permissiveness with respect to possible epenthetic segments. Perhaps unexpectedly, it also allows for greater restrictiveness by way of providing a potential explanation for the general scarcity of default epenthesis systems. The resultant synchronic grammar is instantiated as a set of learned, context-sensitive constraints.

The paper is structured as follows. In the following section a number of ambiguous case studies of epenthesis are presented, and competing analyses are provided. In Section 3, the inherent analytic issues with diagnosing epenthesis are discussed, and a set of relatively theory-neutral diagnostics are introduced. Section 4 presents the entire sample of 50 languages (56 patterns), and an initial broad categorization of the data. In Section 5 quantitative information for the sample is provided along a number of dimensions, and a set of diagnostics are systematically applied. Section 6 provides proposals for a theory of epenthesis, both diachronic and synchronic, and the paper concludes with a general discussion of the findings.

2. The Problem

The correct ranking of constraints for a given grammatical system result, under a psychological model of OT, from a learning process based on evaluation of the linguistic

data (see Tesar & Smolensky 1998). Even though the constraints are typically assumed, the learning task is not trivial, and is complicated by the fact that learners must also acquire the underlying forms of morphemes (cf. Jarosz 2006). If the learning task has been successful, and both constraint rankings and underlying forms are known, then the synchronic formalism provides a clear and explicit computation for all surface forms. The sequence [CVV] is not allowed; neither is [CV] (MAX>>ONSET>>DEP). Nor is the sequence [CV𝕜V] (*k >> *p >> *t). Deterministically, /CVV/ surfaces as [CV𝕥V]. This analysis is proven correct computationally by its ability to produce known surface forms, that is, to reflect the actual words of the language. When those words do not conform to a consistent, unambiguous, and coherent pattern, the correct analysis, and therefore the correct predicted surface forms are not obvious. Consider the following set of problematic languages.

Piggot (1980) describes a process of /t/ epenthesis in Odawa Ojibwa which occurs at the juncture between the vowel-final personal prefixes and vowel-initial non-dependent nouns. In turns out that this process is also accompanied by deletion of unstressed vowels. The actual surface forms that result from the interaction of these proposed rules are given in (1). In (1a), the 2nd person prefix exhibits what might be considered the expected reflex of epenthesis. However, in (1b) the accompanying vowel deletion has rendered the application of /t/ epenthesis opaque. And in (1c) and (1d) deletion occurs because the context for epenthesis is not present (the noun in (1c) belongs to the "dependent" class; the prefix in (1d) does not belong to the personal prefix class).

(1a)	/ki + akatʃi/→[kitakatʃ]	"you are shy"
(1b)	/ni + iʃaː/→[ntiʃaː]	'I go'
(1c)	/ni + oːss/→[noːss]	'my father'

(1d) /kiː + pi + iʃaː + w/→[kiːpiʃaː] ‘he came hither’

In Maori the passive and the nominalizing forms have been described as undergoing /t/ epenthesis, as in (2a) and (2b). However, there is ample evidence for widespread exceptionality (with almost all possible initial segments attested in allomorphs of the passive: –a, -hia,-kia,-lia,-mia,-na,-nia, etc.) (Hale 1973, Elbert & Pukui 1979). The situation is essentially identical in Hawaiian, except that /ʔ/ rather than /t/ is “epenthetic” (see Appendix).

(2a) /mahara + ia/→[maharatia] “remember”
(2b) /mahue + aŋa/→[mahuetaŋa] “leaving”

Furthermore, other verbal forms, such as the causative and reduplicant, and nominal forms, such as the agentive, surface with sequences of vowels as in (2c-e). What might make an epenthesis account even more problematic is that there is no surface alternation evidence to indicate that any morphemes in Maori are consonant-final (see Section 3 for discussion of this issue).

(2c) /whaka + oho/→[whakaoho] “waken”
(2d) /haaere/ + RED→[haaereere] “go”
(2e) /kai + ako/→[kaiako] “teacher”

In Tubatulabal the reduplicant created from vowel-initial stems is described as surfacing with an epenthetic glottal stop (Voegelin 1935), as in (3a).

(3a) aʔaw- (from /aw-/) ‘to gossip’
eʔela- (from /ela-/) ‘to jump’

However, the forms in (3b) and (3c) show vowel coalescence with verbal suffixes, and vowel lengthening in the full reduplicative form (the iterative).

(3b)	/kataxwa + iw + at/→[kataxwiːwat]	'there is dying off'
(3c)	RED + /ela + alat/→[eleːelaːalat]	'he is jumping on both legs like a deer'

Larike has also been described as a /ʔ/ epenthesis system. Laidig (1992) confirms that glottal stop is epenthesized word-initially such that no words surface with initial vowels. /ʔ/ also appears to break up sequences of identical vowels within roots as in (4a).

(4a)	/waa/→[waʔa]	"root"

However, epenthesis fails to occur across morphological boundaries when the vowels are different (4b), and deletion occurs when the two vowels are identical (4c).

(4b)	/bebeku + i/→[bebekui]	"duck pl"
(4c)	/riʔa + aʔu/→[riʔaʔu]	"for me"

Ebert (1996) describes the following pattern in Kodava: /j/ epenthesis after front vowels (5a); /v/ epenthesis after back vowels (5b); /k/ epenthesis before /a/ (5c); and deletion when the two vowels are identical (5d).

(5a)	/elli + uu/ →[ellijuu]	'wherever'
(5b)	/boɳɖu + aa/→[boɳɖuvaa]	'is it necessary?'
(5c)	/kuɖi + a/ →[kuɖika]	'let's drink!'
(5d)	/ʌɭɯdi + ija/→[ʌɭɯdija]	'you wrote'

/k/ epenthesis, however, also occurs when the stem ends in an /n/, as in (5e) (accompanied by voicing).

(5e) /tin + adɯ/ → [tingadɯ] "let him eat"

Analysis of the foregoing set of languages is problematic due to the complexity of the observed patterns. The problem can be stated in the following way: as a choice between an epenthesis analysis, a deletion analysis, or a 'morphological analysis' (Hale 1973). If one has compelling reason to decide that some or all of the patterns described above are not actually epenthesis, but merely 'epenthesis-like', then the formal system for describing epenthesis is not responsible for computing them. This decision then, clearly, is a critical one for the questions of this paper.

Unfortunately, there exists no formal mechanism within phonological theory to definitively decide the issue. A framework of interacting constraints of the right type is quite capable of producing a pattern that differentiates vowel hiatus environments created by prefixation from those under suffixation: deleting in the former, epenthesizing in the latter, for example; or deleting a suffix-initial vowel to preserve maximal word limits, single footed prosodic words, and left-aligned bimoraic feet, except where a sequence of identical vowels would result, and metathesizing to avoid [ni] sequences (de Lacy 2003). Such a framework does not require a single repair to hold in *all* possible environments. However, determining how much to relativize such constraints is not a formal part of the theory. Nor is determining how to split the analytic work between representations and constraints.

Consider two competing analyses of the language patterns in (1) through (5) by authors who both work within an OT framework (de Lacy 2006 and Lombardi 2002). Table 1 indicates whether the given pattern is assessed as representing genuine epenthesis (✔), or belonging only to the class of "apparent epenthesis phenomena" (✗). Languages

in the unfilled cells are ones the authors have explicitly argued for or against. The gray filled cells represent my own application of their diagnostics.

Table 1
Two differing diagnostics for epenthesis: DL (de Lacy 2006); L (Lombardi 2002).

Language	**DL**	**L**	**Disqualification**
Odawa	✔	✗	Morphologically restricted
Maori	✔	✗	Not 'elsewhere' case
Tubatulabal	✔	✗	Not 'elsewhere' case.
Larike	✔	✔	
Kodava	✗	✗	Restricted distribution, non-optimal; Not 'elsewhere' case

As can be seen from Table 1, the two diagnostics agree only in accepting epenthesis in Larike, and rejecting epenthesis in Kodava. In all instances, however, they share similar reasons for rejecting certain patterns which have to do with lack of generality. Where they disagree, their designations of what is too restricted in scope to qualify for epenthesis differ.

3. "True Epenthesis"

The goal of this section will be to establish clear, consistent, and pragmatic conditions for identifying epenthesis patterns – patterns that fall within the scope of study, and *must* be accounted for by any linguistic theory of epenthesis. The first step will be to strictly limit that scope. Therefore, only patterns of epenthesis applying in situations which would otherwise result in adjacent vowels will be considered. Additionally, the epenthetic material must consist of a single consonantal segment, so epenthesis of a CV syllable, for example to satisfy a template or minimal-word requirement, will be excluded from consideration. Furthermore, this epenthetic segment cannot be analyzed as belonging to the underlying structure of any of the morphemes involved. In other words, there should be no segment in the input than can be identified as a correspondent to the epenthetic

segment in the output. Thus, cases that can be analyzed as glide formation in which an underlying vowel surfaces as a homorganic glide under affixation will also be rejected. An example of this from Ilokano is given in (6).

(6) /pag + santo + an/ → [pagsantwan] 'to make into a saint'

All these phenomena are excluded by stipulation, not from the purview of the theory as a whole, but of the subset of the model concerned with 'epenthesis' in a distilled and simplified form, about which theoretical universals of typology can be straight-forwardly extracted.

Establishing diagnostics to separate on the one hand "true" or "default" epenthesis patterns, and on the other a whole host of non-epenthesis phenomena will be more controversial. Therefore, we will proceed in a conservative way, and begin with those requirements which are the closest to theory-neutral – more properly, neutral with respect to the particular claims under investigation. It is critical to the arguments of generative theory, as well as those in this paper, that the pattern under discussion be a synchronic one. This is to say that the linguistic data be convincingly analyzable as the product of an active grammar, and not some other mechanism. While an argument can certainly be made for expecting possible sound changes to be restricted in the same way as possible grammars, this is by no means an accepted conclusion. To avoid the complications arising from these cases, potential epenthesis patterns which are clearly restricted to the diachronic realm will be excluded from analysis.

An example of this is Maru, which diverged from other Kachin languages by developing stops word-finally, the place of articulation determined by vowel quality (Burling 1966). Epenthesis, if the process can be described as such, occurred at a

particular point in time; it is not observable in synchronic alternations. In other instances it is not so easy to establish the status of a given pattern. Because static patterns – that is, phonotactic observations over lexical items – may easily be the residue of historic processes, we will require that the candidate epenthesis pattern be realized in morphological environments. Furthermore, as a requirement of transparency, the morphology must show a surface alternation which can provide evidence that the segment in question is not underlying. An example of the required type of alternation is given in (7), where the genitive morpheme in Tubatulabal is shown suffixed to a vowel-final stem in (7a), and to a consonant-final stem in (7b).

(7a) /kʊmuː + in/→[kʊmuʔʊn] "of his own father"
(7b) /pʊnihw + in/→[pʊnihwin] "of his own skunk"

The fact that the suffix surfaces intact in (7b) provides evidence for its given underlying form (vowel harmony applies in (7a) but is irrelevant to the argument here). The evidence in (7) is not enough to establish conclusively that an epenthesis process is at work, but it is minimally necessary, for our purposes, to establish that such a process *might* be active in the synchronic grammar. In the following sections additional conditions for eliminating other possible analyses will be established.

3a The Coincidence Condition: Epenthesis or Deletion?

Axininca Campa is a popular textbook example of epenthesis, and provides a nice exemplification of the implicit evidence employed in phonological analysis. The language is standardly described as exhibiting both /t/ epenthesis (to break up vowel

clusters), and /a/ epenthesis (to break up consonant clusters) (Spring 1990). There is, however, at least one viable alternative analysis, shown in row A of Table 2.

Table 2
Four possible analyses of Axininca Campa, following Spring (1990).

	Analysis	Underlying Representations and Rules
A	/a/ deletion	/i/ + /n/ + /kima/ + /piro/→[iŋkimapiro]
		/i/ + /n/ + /kima/ + /i/→[iŋkimi]
	/t/ deletion	/pijat/ + /i/→[pijati]
		/pijat/ + /piro/→[pijapiro]
B	/a/ epenthesis	/i/ + /n/ + /kim/ + /piro/→ [iŋkimapiro]
		/i/ + /n/ + /kim/ + /i/→ [iŋkimi]
	/t/ epenthesis	/pija/ + /i/→ [pijati]
		/pija/ + /piro/→ [pijapiro]

Analyses (A) and (B) do equally well in consistency of application of their respective rules (either deletion of both segments, or epenthesis of both segments). And, taken individually, there is nothing to choose between in selecting the underlying forms for the morphemes. It is only when the URs are considered in the aggregate that a point of distinction emerges.

It can be shown that analysis (A) requires one of two outcomes for the underlying forms. Either all verbal stems end only in /a/ or /t/, or all verbal suffixes begin only with /a/ or /t/. In either case a fairly large number of morphemes must adhere to a very constrained template (only two phonemes out of an inventory of 20 consonants and 3 vowels). The choice of analysis then hinges on whether or not this is considered too much of a coincidence. Again, there seems to be no formal treatment of this aspect of phonological analysis, although various authors have proposed intuitive solutions[2].

[2] Spring (1990, p.53) writes of Axininca Campa: "The result…would be that while stems begin and end in c's and v's of various qualities, all verbal suffixes would begin with /a/ or /t/, a clearly undesirable result." For Payne (1981:56-57)

The issue of the appropriate choice in synchronic analysis between deletion and epenthesis is one that goes well beyond the apparently small scope of the original question of this paper. The choice is relevant not only to the analysis of the linguist, but to the analysis of the learner. More generally, the question is how acquirers of language learn underlying representations when all they are given are surface forms. This unresolved problem is central to learning theory within linguistics (see, e.g., Tesar and Smolensky 1998, Jarosz 2006), and it will not be solved here. However, it will be argued that a model of this learning mechanism is ultimately necessary to account for the typology of epenthesis. Furthermore, although formal linguistic theory may omit reference to the number of morphemes involved, such factors can be seen to play a critical if covert role in analysis. That covert material will be explicity assessed in the following development of quantitative measures for separating out "apparent" epenthesis patterns that ought instead to be analyzed as deletion.

3b The Benchmark: Axininca Campa v. Buryat

The morphophonology of Axininca Campa has been previously analyzed by different authors, with varying conclusions (McCarthy & Prince 1994, Lombardi 2002, de Lacy 2006). Not all of these treatments present, or include within their analysis, exactly the

---

> "The epenthesis solution…avoids the arbitrariness of evaluating between the two deletion solutions. Second, and more important, it explanatorily accounts for the nature of the segments that must be either epenthesized or deleted by rule…there seems to me to be no natural reason why a /t/ and an /a/ would be deleted rather than any other consonant or vowel…on the other hand, …if a vowel is epenthesized to break up a consonant cluster, it would be natural for that vowel to be the unmarked /a/."

Kenstowicz and Kisseberth (1979:87) identify "[a] very general methodological principle". Namely, "if there is a 'pathological' phonetic pattern (i.e., a pattern that disobeys general principles of the language…)…then it is likely that this pattern is the consequence of a rule, rather than a property of the underlying form of the language (on the other hand, advocates of morpheme structure constraints might well argue that just such a rule should apply at the representational level. See, e.g., Kaye 1974).

same set of data. Since this paper is remaining as agnostic as possible about the theoretical status of epenthesis at this stage, the focus is on making available all *potentially* relevant data for consideration. This full set is given in Table 3 (as well as the Appendix). The underlying representations assumed for the analysis of /a/ epentheis and /t/ epenthesis (row B of Table 2) are used, but contradictory alternations are provided as well. From this point forward, the minimal evidence for the alternation (*a la* the example in (7)) will be omitted; it may be assumed. All positive evidence (morphemes that can be counted towards an epenthesis analysis) will be presented on the left-hand side of the table, with the assumed URs; all negative evidence, on the right (this is the organization of the evidence in the Appendix as well). /t/ epenthesis has been claimed by both Payne (1981) and Spring (1990) to be a general process. But since exceptions were found, all affixes are verified, and omitted if the minimally necessary alternation could not be found within the sources.

Table 3
Axininca Campa Epenthesis Data (Payne 1981, Spring 1990)

Axininca Campa [Payne 1981]				
t-Epenthesis		**Other**		
Verbal Suffixes				
Reflexive		**Interruptive**		(ts
/hoti + a/→[hotita]	"he got in (put himself in)"	/pi + in + koma + ima + i/→ [piŋkomat͡simati]	"(you) ready to paddle"	
Perfect		**Non-Future**		
/no + na + ak + i + ro/→ [nonatakiro]	"I have carried it"	/no + pisi/ + /i/ → [nopisi͡tsi]	"I swept"	
Dative		**Distributive**		Deletion
/in + koma + ako/→[iŋkomatako]	"he paddles for"	/ir + impoi + i͡ts^{h}i + an + ak + a/→ [himpoi͡ts^{h}itanaka]	"he followed along behind, in addition, departing"	
Repetitive				
/in + koma + aa/→ [iŋkomataa]	"he paddles again"			
Future		**Progressive + Future**		
/in + koma + i/ →[iŋkomati]	"he will paddle"	/hoti + a + i/→[hotitatʃi]	"he will be putting in"	tʃ
Infinitive		**Reflexive Future**		
/koma + aants^{h}i/→[komataants^{h}i]	"to paddle"	/ir + n + kisi + ia/→[iŋkisitʃa]	"he will comb himself"	
Receiving		Also [ija] [iija] [ja]		
/ir + kaɴt + βi + aβ + auɻ + a + ri/→ [ikaɴtaβitaβaari]	"he said to the one arriving"	Nominal Suffixes		
		Diminutive		
		/hito + iriki/ → [hitoiriki]	"little spiders"	
Resolved		/sampaa + iriki/→ [sampairiki]	"little balsas"	Fusion
/aa + auɻ/→[aatauɻ]	"he will go back"	**'Drink'**		
Passive		/siŋki + a/ → [siŋkitʃa]	"corn drink"	tʃ
/ir + ii + ai + i + ri/→[hiitai͡tsiri]	"that which is named"	/kimi + a/ → [kimija]	"squash drink"	j
Purpose		Verbal/Nominal Prefixes		
/no + ojaa + asi + a + ri/→ [nojaatasitari]	"I followed (for the purpose of visiting)"	**1st person singular**		Deletion
		/no + ir + i/→[niri]	"I will drink"	
Reason		**2nd person singular**		
/βi + t^{h}a + ant + ap + auɻ + ak + a/→ [βit^{h}atantapaaka]	"therefore he greeted on arrival"	/pi + iit^{h}oŋk + i/→[piit^{h}oŋki]	"you will climb"	
Time: Early		**3rd person singular female**		
/o + ɲaa + aman + i/→ [oɲaatamani]	"it dawned (early)"	/o + aa͡tsik + i/→[aa͡tsiki]	"she will step"	
Reciprocal		**1st person plural**		
/arii + aβ + ak + auɻ + ia/→	"having arrived to each other	/a + ook + i/→[ooki]	"we will abandon"	

[ariitaβakaija]	(visited)"			
Modal Plural				
/ha + aij + an + aɯ + i + ni/→ [hataijanaini]	"they went back, departing"			

/t/ epenthesis has been described as applying to verbal suffixes. 14 such verbal suffixes were verified[3]. However, there are a handful of verbal suffixes that are exceptions to this pattern. The Non-Future, Interruptive, Progressive, and Reflexive Future either show epenthesis of an affricate (tʃ or ts), or a pattern of multiple alternates. There is also an irregularity in the otherwise regular epenthesis pattern for the Future. When combined with the Progressive, /tʃ/, rather than /t/ is epenthesized (see Table 3). There are a total of 4 exceptions, and 1 irregularity to the /t/ epenthesis pattern among the verbal suffixes.

Table 3 also shows that for suffixes in the nominal paradigm vowel hiatus is tolerated, or repaired via coalescence (-/iriki/), or via epenthesis of /tʃ/ or /j/ (-/a/). Nor is there any evidence for /t/ epenthesis in reduplication, where the person prefix /n/- is repeated as necessary to avoid an onsetless syllable. And in the domain of prefixation, avoidance of VV sequences primarily involves deletion of the first vowel of the sequence. To differentiate these contexts from the more restricted domain of verbal suffixation, let us define a class of morphemes that could, prosodically speaking, exhibit epenthesis behavior, but don't. These are Non-Participating Morphemes. Taking all quantitative data from Table 3 gives a final tally for Axininca Campa at 14 participating

[3] There are a handful of other verbal suffixes for which the relevant forms are not readily found in Payne (1981): Arrival:-/ap/, Departure:-/an/, There/And/Back:-/ak/, 1st person incl. obj:-/ai/, Causative:-/ak/, Rapid:-/apaint/.

verbal suffixes, 4 exceptions, 1 irregularity, and 6 non-participating morphemes (2 nominal suffixes; and 4 verbal prefixes).

In order to determine the relative size of these measurements, let us next compare Axininca Campa to another proposed epenthesis language. Buryat has been described as exhibiting /g/ epenthesis, and the evidence for and against this analysis has been culled from Poppe (1960) and presented in Table 4. The pattern is complex but phonologically conditioned. In situations of vowel hiatus two different dominant outcomes arise: when the first vowel is short, it deletes; and when the first vowel is long, /g/ is epenthesized. It should also be noted that a general phonetic process applies in Buryat which causes the epenthetic segment to surface in the following forms: [g] before front vowels; the uvular fricative [ʁ] between back vowels, and the uvular stop [ɢ] after front and before back vowels. [k] is also described as an allophone of /g/ in other contexts. A separate voiceless aspirated velar stop is listed as a marginal phoneme that only occurs in loanwords. Vowel harmony applies, altering the forms of most suffixes.

Table 4
Buryat epenthesis pattern (Poppe 1960)

Buryat [Poppe 1960]				
g Epenthesis		Other		
V:+V…				
Nominal Suffixes				
'of/belonging to'				Suppletion
/ʃere: + ai/→[ʃere:gei]	"of the table"	/n/: after /i:/, diphthongs; /i:n/: after single vowel		
Reflexive Possessive attached to oblique stem				
/taxa: + a:/→[taxa:ga:]	"own hen"	/ja:/: after single V or /n/		
'along, over' with reference to motion		**Direct Object**		
/sa: + u:r/→[sa:gu:r]	"over there"	/zygi: + i:ji/→[zygi:ji]	"the bee" D.O.	
Adverb from noun		**Reflexive Possessive with pure-relational suffixes**		
/yglø: + y:r/→[yglø:gy:r]	"in the morning"	/a:/˛/ga:/˛/ja:/˛/nga:/˛/n/		
2nd verbal noun of the present		**Plural**		
/bai + aaša + n^{j}/→[baigaašan^{j}]	"The one who stands"	/d/˛/u:d/˛/nu:d/˛/nar/		
'by means of'				
/noxoi + a:r/→[noxoigo:r]	"by means of the dog"			
Verbal Suffixes				
Perfective gerund		**Causative**		Suppletion
/u: + a:d/→[u:ga:d]	"having drunk"	/ga/˛/gaa/˛/uul/˛/lga/˛/aa/		
Present imperative				
/xyle: + i: + ʃ/→[xyle:gi:ʃ]	"wait indeed!" 2 sg			
Future				
/xyle: + u:ʒa + b/→[xyle:gy:ʒeb]	"I shall wait"			
Future imperative				
/xyle: + a:rai + ʃ/→[xyle:ge:reiʃ]	"wait!" (later) 2sg			
Verbal noun of the imperfect				
/bai + a:/→[baiga:]	"he was"			
		V+V…		
		Deletion		

The epenthesis pattern is observed for certain nominal suffixes, such as the Instrumental. However, there are other morphemes for which suppletion occurs in this context (such as the Plural), and others that show a pattern of deletion (such as the Direct Object). The balance of evidence in the nominal domain is not entirely clear since for the

majority of suffixes no long vowel-final stem inflections are provided[4]. In verbal suffixes, however, the epenthesis pattern is almost completely exceptionless, with only the Causative failing to conform[5].

The data in Table 4 show, for all long vowel contexts, a total number of 11 participating morphemes (6 nominal suffixes; 5 verbal suffixes); 4 non-participating (3 nominal; 1 verbal); and 2 irregular forms (nominal). Axininca Campa has more participating morphemes than Buryat, but it also has more non-participating forms. A simple count threshold of participating morphemes will bias a diagnostic towards systems with richer morphological systems, or even morphological systems for which there is more comprehensive available data. One way to avoid such a bias is to take a ratio as the relevant measure, a ratio, for example, of participating morphemes to total number of morphemes. This will act to normalize the languages to be compared to a certain degree. The ratio measure gives the following scores: 14/15 ~ 56% for Axininca Campa, and 11/17 ~ 65% for Buryat.

Under the ratio measure proposed above, Buryat emerges as a more robust example of true epenthesis than Axininca Campa. This means that if Axininca Campa passes the epenthesis diagnostic, then so must Buryat. This is not, however, the final word. There are further complications in comparing these two languages. We may decide that non-participating morphemes should be separated out from exceptions, or even that non-participating morphemes are irrelevant to the assessment of epenthesis behavior. If

[4] - /uːʃan/ 'of occupation or indication of social group'; -/aːling/,-/aːrxin/,'collectives'; -/aːxai/,-/uːl/,'suffix of the actor'; -/aːri/,-/uːr/, 'suffixes of the instrument'; -/uːri/, 'suffix of place of action; -/aːl/,-/aːn/, 'suffixes of the act'; -/aː/,-/aːrga/,-/aːdahan/-/aːhan/, 'suffixes of the result and object of action';-/aːd/, 'distributive (for numerals); -/uː/,-/uːxai/, 'nominalizes adjectives'; -/aːr/,'adverb of manner (from adjectives)'

[5] No alternations are provided for the 2nd person verbal nominalizer -/aːʃa/, or the passive -/aːtai/.

epenthesis must only be assessed for generality with respect to a particular phonological domain, then the desired measure is the ratio between participating morphemes and all morphemes within the relevant domain. However, there is no automatic procedure either for defining a possible phonological domain, or determining the correct domain of application (see Kenstowicz & Kisseberth 1977 Section 2.2 for discussion). For Axininca Campa, 14 verbal suffixes exhibit the epenthetic behavior, while 4 consistently fail to do so, as well as 1 particular suffix combination: 14/19 ~ 74%. The appropriate phonological domain in Buryat is somewhat more difficult to ascertain. For long-vowel suffixation, the number remains 11/17, or 65%. For verbal suffixation in long-vowel environments alone, the score is 5/6, or 83%. It is clearly critical exactly how the diagnostic measure is defined. Under certain interpretations we find that Axininca Campa provides better evidence than Buryat, but under others we find the opposite result.[6]

4. Typology Revisited

In attempting to discover an appropriate quantitative measure of the linguistic data a set of statistics was considered: 1) Counts of the number of morphemes that exhibit epenthesis behavior within a phonological domain, 2) the number of exceptions within that domain, 3) the number of irregular lexical items, 4) the total number of morphemes that meet general prosodic requirement but fail to participate, and finally a ratio of participating morphemes to total morphemes, 5) over the domain, and 6) over the language. While the preferred diagnostic will depend on other theoretical commitments,

[6] All numbers provided are conservative estimates. A given suffix is only listed if the relevant surface forms are actually transcribed in the source. The most liberal estimate, including the 16 undetermined suffixes in Buryat, and the 6 in Axininca Campa, would place the number of participating morphemes as 20 in Axininca Campa, and 27 in Buryat.

we will assume that 1)-6) provide comprehensive coverage of the evidence that is relevant to the issue. Based on this assumption, these measures are now applied to a larger sample of languages: a total of 56 candidate epenthesis patterns, collected and systematically analyzed in accordance with the criteria established in Section 3.

The sample was constructed in the following way. The references cited in de Lacy (2006), Lombardi (2002), and Vaux & Hall (2001) were reviewed and augmented as necessary. In most cases, at least one grammar was found with sufficient detail to assess the robustness of the proposed pattern. Table 5 indicates the range of epenthetic segments that have been proposed by the above authors (and a few others). The languages analyzed in this paper are indicated by underlining. Not every pattern was investigated, but every effort was made to include each proposed case of /t/ and /k/ epenthesis, in order to provide the most data for the markedness claim under investigation.

Additionally, a search was performed on the P-base database (Mielke 2007) using the keywords "hiatus" and "insert". In total, about 130 languages were returned; of these, a large number actually involved vowel epenthesis, another subset were not clearly prosodically conditioned from their descriptions. Others of the languages were already present in the sample. Of the remainder, the vast majority were listed as involving epenthesis of one or more of the set of {j,w,v,h,ʔ}. In total, only 10 languages were listed as involving other segments: {r,d,fricative,n,l,m}. Of the total set of consonant epenthesis candidates, 7 were investigated; these were selected at random, and happened to all involve glide epenthesis. Finally, additional languages were acquired fortuitously as the author encountered citations in various sources.

Table 5
Proposed Consonant Epenthesis Languages by segment. Underlined languages were investigated by the author and appear in the Appendix.

Seg	Language	Seg	Language
t	Axininca Campa, Maori, Odawa Ojibwa, French, Amharic, Plains Cree, Maru, Finnish, Korean, Kodava	j	Turkish, Uyghur, Greenlandic, various Indic languages, Arabic, Slavic, Tamil, Kodava
k	Maru,Kodava	h	Ayutla Mixtec, Chipewyan, Huariapano, Slave (Bear Lake, Hare), Tigre, Tucanoan, Yagua,Yucatec Maya, Huaripano, Onondaga
g	Mongolian; Buryat	w	Abajero Guajiro, Greenlandic, Arabic, Chamicuro, Tamil
r	English, German, Uyghur, Zaraitzu Basque, Seville Spanish, Anejom, Japanese, Southern Tati	ʔ	Chadic, Cupeno, Larike, Misantla Totonac, Mohawk, Tsishaath Nootka, Hawaiian, Arabic, Selayarese, German, Ilokano, Czech, Kisar, Malay, Koryak, Indonesian, Gokana, English, Konni,Tunica, Tubatulabal, Nancowry, Tamil
n	Korean, Greek, Dutch, German dialects, Sanskrit, Murut, Tunica	x	Land Dayak
l	Bristol English, Midlands American English, Motu	ʃ	Basque dialects
v	Marathi	ʒ	Cretan and Mani Greek, Basque dialects
b	Basque dialects	ŋ	Buginese
s/z	French, Land Dayak, Dominican Spanish	ɴ	Inuktitut, East Greenlandic, Uradhi, Kaingang

It is important to note that this sample is very much *not* a random sample of languages. It is a sample designed to find the maximum number of true epenthesis patterns by investigating only languages where previous researchers have claimed to observe epenthesis patterns. However, in some cases very little data were available from the cited source (and sometimes, even from additional sources). The full set of data used for analysis for each of the languages in the following tables is available in the appendix (in alphabetical order by language). In (8) are listed a number of criteria on which the languages of the sample are assessed. These languages have been divided into several groupings based on these criteria. These are shown in Tables 6-9.

(8) C1. Prosodic conditioning (optimal surface forms)
C2. Repair to satisfy ONSET
C3. Epenthesis repair only for onsetless syllables
C4. Exceptionless pattern
C5. Single epenthetic segment (default)

The requirements for prosodic conditioning, specifically to satisfy a hypothesized onset constraint (alternatively, to break up vowel sequences), have been discussed above. C1, C2, satisfaction of the alternation evidence in (7), and morphological conditioning are considered minimally necessary conditions for consideration as an epenthesis pattern. That several languages of the sample fail to meet one or more of these pre-conditions can be seen in Table 6.

Table 6
Languages of the sample that fail one or more of the epenthesis pre-conditions

	Seg.	**Language**	**Criterion failure**	**Source(s)**
1	k	Wolof	C1,C2	Ka (1994)
2	k	Danish	(7)	Glyn Jones & Gade (1981) Herslund (2002)
3	t	Odawa	C1	Piggot (1980)
4	t	Maru	C1,C2	Burling (1966)
5	t	Amharic	C2	Broselow(1984)
6	r	Uyghur	C1	Hahn (1991) Hahn (1992)
7	v/w,j	Marathi	(7)	Wali (2005)
8	t,n	Korean	C1, C2	Renaud (1974) Lee (1998)
9	ʔ	Larike	Non- morphologically conditioned	Laidig (1992)
10	ʔ	Cairene Arabic	Non- morphologically conditioned	Watson (2002)
11	k	Kodava	C2	Ebert (1996)
12	j,w,ʔ	Tamil	Non- morphologically conditioned	Christdas (1988)

Almost all languages show some degree of exceptionality, or irregularity, within even a fairly robust epenthesis pattern (C4). It has already been shown that certain morphemes in Axininca Campa make up part of a suppletive allomorphy pattern, for example. And, in Turkish (pattern #53; Table 9), one of the most consistent epenthesis languages, the

Genitive sometimes undergoes /j/ epenthesis, and sometimes surfaces with a 'ghost' /s/ or /n/ (with optional final /n/). See (9). Note that Turkish requires agreement on several vowel features within a word, accounting for the alternations on the vowel of the Genitive suffix.

(9)

Stem	Genitive	Gloss
baba	babasɨ	'father'
su	sujun	'water'
el	eli	'hand'

Table 9 also reveals that the vast majority of the sample languages make use of multiple repairs to avoid onsetless syllables (C3). This is clearly true in Axininca Campa where only the suffixes exhibit the epenthesis alternation, and one of the two adjacent vowels typically deletes under prefixation. The opposite patterning holds for Plains Cree (Pattern #38; Table 9), where /t/ epenthesis occurs with (most) prefixes, but epenthesis of /j/, /w/, deletion of one of the adjacent vowels, or coalescence occurs under suffixation. See (10).

(10)

/wajawi: + a:mo/	→[wajawi:ja:mo]	'run outside (flee)'
/ni:pa: + ohte:/	→[ni:pa:hte:w]	'He walks in the dark'
/ko:na + ehk/	→[ko:nihk]	'in the snow'
/o + te:m + iji + a/	→[ote:mijiwa]	'his horse'

Lack of uniformity is the norm, rather then the exception, and it is particularly true when comparing word-initial and morpheme-boundary environments. In fact, only one language seems to epenthesize in both cases, as well as epenthesize the same segment, Misantla Totonac: /ʔ/ (Pattern #52; Table 9). Similarly, it is rarely the case that

the occasional vowel-initial word or syllable does not arise. Vowel-initial words appear quite regularly in Axininca Campa, for example, and word-internal vowel sequences are sometimes observed as well. These sequences result from a process of intervocalic glide deletion. Examples of this can be seen in the data in Table 3 when /ɯ/ appears underlyingly. The vocative forms, as well as a closely related derivational kinship form, surface in somewhat unpredictable ways, sometimes containing sequences of adjacent vowels: contrast the 1st person: [nijoti], with the vocative form: [ijoini] of "mother-in-law (male ego)". Given that almost all languages studied failed to pass either C3 or C4, those criteria will no longer be used in their absolute forms.

Firstly, take languages that fail criterion C5 in (8): no single default epenthetic consonant. This subset of languages is given in Tables 7 and 8; all these languages can be analyzed as employing epenthesis of multiple segments. In Table 7, those segments are phonologically unpredictable. The default analysis for these languages is stipulated as deletion rather than epenthesis, for consistency with the coincidence conditions of Section 3a. This set of patterns is contrasted with those of Table 8 in which the segments are conditioned by phonological context. In Table 8, all languages can be characterized as examples of "assimilative" epenthesis[7].

[7] The exception to this is West Greenlandic, which is perhaps best described as dissimilative epenthesis: the place of the glide disagrees with the place of the surrounding vowels.

Table 7
Languages of the sample that fail C5, segments unpredictable

	Seg.	Language	Phonological Domain	Source(s)
13	t,z,p,n,r,l, g,d	French	Word boundaries	Tranel (1981)
14	dj,s,k	Finnish	Transitive verbal loans	Luthy (1973) Sulkala & Karjalainen (1992)
15	ʔ,h,k,l,m	Hawaiian	Passive	Elbert & Pukui (1979)
16	j,w,t	Abujhmaria	Suffixation*	Natarajan (1985)
17	w,j	Sinhalese	Suffixes belonging to certain declension types*	Geiger (1938)

What can also be noted about the set of languages in Table 8 is the fact that only a very small set of segments is involved in this type of epenthesis system: w, v, or ʊ (sometimes h) – in back vowel contexts – , j – in front vowel contexts –, and sometimes ʔ – usually when the two vowels are identical. Take the pattern in Malay, illustrated in (11). Under suffixation, when the stem-final vowel is /u/, /w/ is epenthesized (11a); when the stem-final vowel is /i/, /j/ is epenthesized (11b); and when the stem-final vowel is non-high, or the two vowels are identical, /ʔ/ is epenthesized (14c) (Onn 1976).

(11a) /bantu + an/→[bantuwan] 'relief'
(11b) /udʒi + an/→[udʒijan] /test/
(11c) /məŋ + gula + i/→[məŋgulaʔi] 'cause to sweeten'

Although such patterns could be described as involving a single underlying segment – namely, a glide unspecified for place – the difference among the surface realizations is

* Partially predictable by vowel. See Appendix.

quite significant, both articulation-wise, as well as auditorily. For this reason, this outcome will be characterized as involving phonological assimilation. This will be contrasted with a case of phonetic assimilation to be discussed below.

Table 8
Languages of the sample that fail C5, segment predictable

	Seg.	Language	Phonological Domain	Source(s)
18	j,w,ʔ	Malay	Suffixation	Onn (1976)
19	j,w,ʔ	Wolof	Suffixation	Ka (1994)
20	j,w,ʔ	Guinaang	Suffixation	Gieser (1969)
21	j,w,ʔ	Karo Batak	V+V†	Woolams (1996)
22	j,w,ʔ	Hausa	Certain suffixes	Jagger (2001)
23	w,j	Balangao	V+V	Shetler (1976)
24	w,j	Dakota	V+V	Shaw (1980)
25	w,j	Ao	V+V⁘	Gowda (1975)
26	w,j	Manipuri	V+V	Bhat & Ningomba (1997)
27	w,j	Argobba	Suffixation⁘	Leslau (1997)
28	w,j	Alywarra	word boundaries†	Yallop (1977)
29	v,j	West Greenlandic	Prefixation involving three or more vowels†⁘	Rischel (1974)
30	v,j	Kodava	Suffixation	Ebert (1996)
31	ʋ,j	Malayalam	Suffixation	Asher & Kumari (1997)
32	j,ʔ	Ilokano	Prefixes	Hayes & Abadi (1989)
33	w,j, h	Cairene Arabic	Heterogeneous collection of templates and suffixes	Watson (2002)
34	ʋ,j	Dutch	Host+Clitic	Booij (1995)

Finally, take systems that are best analyzed as epenthesis of a single underlying segment. The languages shown in Table 9 pass the preliminary diagnostics C1 and C2, as well as C5. Axininca Campa and Buryat, discussed extensively in previous sections, are

† Optional in all environments.
⁘ Not completely predictable by vowel. See Appendix.

included here. Although C3 and C4 were rejected in their absolute forms, they will be resurrected here in order to make a final determination with respect to robustness of the epenthesis pattern. This will follow directly from the discussion in Section 3.

In Table 9, the phonological domain of epenthesis is listed, followed by the count of the number of Participating Morphemes (PM) in that domain. The total number of Non-participating Morphemes – in any domain that meets the prosodic requirements – is listed in the next column (NP). Finally, Exceptions (Ex) are defined as those morphemes that fall within the phonological domain but fail to participate (as defined previously, and as distinct from lexically irregular items)[8]. The Exceptions will always be a subset of the Non-Participating Morphemes. This type of accounting will allow us to approach a quantitative method for diagnosing default epenthesis patterns. However, a note on the method of counting is first necessary.

[8] The 'irregulars' from the discussion of Axininca Campa and Buryat will be folded in to the class of Exceptions. This is because the irregular behavior is not limited to a handful of lexical items, but to a large set of stems to which the morpheme attaches (Buryat), or an entire set of trimorphemic forms (Axininca Campa).

Table 9
Languages of the sample that pass C5

	Seg.	Language	Phonological Domain	PM	NP	Ex	Source(s)
35	k	Dakota	3rd personal plural	1	5	0	Shaw (1980)
36	t	Maori	Trimoraic roots of light syllables in the Passive & Gerund*	2	7	0	Hale (1973) Bauer (1993) De Lacy (2003)
37	r	S. Tati		1	3		Yar-Shater (1969)
38	t	Cree	personal possessive prefixes	14	14	2	Wolfart (1973)
39	n	Dutch	ə-final host+clitic†	17	3	0	Booij (1995)
40	k	Waropen	verbal person prefixes	7	5	2	Anceaux (1961) Held (1942)
41	g	Buryat	Verbal Suffixes and long-vowel final stems	5	7	1	Poppe (1960)
42	t	A. Campa	Verbal Suffixes	14	11	5	Payne (1981)
43	ɣ	Greek	Verbal Suffixes	2	7	1	Newton (1972) Malikouti-Drachman (2009) Joseph & Philippaki-Warburton (1987)
44	w	Chamicuro	Prefixes	1	4	3	Parker (1989)
45	j	SE Armenian	1-syllable words	4	6	0	Vaux (1998)
46	j	Uyghur	verbal suffixes	1	5	5	Hahn (1991) Hahn (1992)
47	ʔ	Nancowry	Nominal affixation	1	5	3	Radhakrishnan (1981)
48	ʔ	Tsishaath Nootka	V+V	2	4	4	Stonham (1999)
49	ʔ	Cupeño	V+V	3	5	5	Crowhurst (1994) Hill (2005)
50	ʔ	Selayarese	V+V;V=V	5	1	0	Mithun (1986)
51	ʔ	Tubatulabal	V+V	7	5+	5+	Voegelin (1935)
52	ʔ	Misantla Totonac	V+V	15+	9+	9+	Mackay (1994, 1999)
53	j	Turkish	Suffixation	21	10	10	Underhill (1976)
54	j	Berber	Suffixation	7	2	2	Guerssel (1986) Hdouch (2008) Jilali (1976)
55	ʔ	Tamil	compounds	1	0	0	Christdas (1988)
56	ʔ	Malay	prefixation	2	0	0	Onn (1976)

* Adopting the analysis of de Lacy (2003)
† Optional variant of three possible repairs

The issue of problematic counting contexts surfaced at the end of Section 3. Further complications are discussed here. Epenthesis might be observed among the personal prefixes, but only with dependent stems. Non-dependent stems are thus exceptions to this rule. Similarly, epenthesis might occur in compounding, but not in reduplication. Thus, reduplication is a non-participating process. Epenthesis might occur in word-initial position, or it might fail to occur in exactly that environment. Epenthesis might occur when the first vowel in the sequence is front, but not when that vowel is back. This type of complexity is precisely what one might expect of natural language, but it does complicate quantitative analysis. It also calls for the more explicit definition of a number of constructs within phonological analysis. Pending that difficult work, we will proceed with the crude method of counting each of the listed instances as one 'morpheme', participating, non-participating, or exceptional (if inside the given phonological domain).

5. Possible Diagnostics

Overly strict epenthesis diagnostics – ones that are completely intolerant of irregularity or variance – will rule out languages that are commonly held to be particularly good (even archetypal) examples of epenthesis, such as Axininca Campa. Such standards have the potential to eliminate the possibility of an epenthetic analysis for just about every candidate language in our sample. Backing off from an expectation of completely exceptionless behavior requires the analyst to impose a clear threshold for diagnosing default epenthesis systems. This threshold must be defined with respect to a particular criterion, or set of criteria, and with respect to a particular way of counting.

The counting method for the numbers in Table 9 has already been discussed[9]. The three measures that will contribute to the evaluation score are the three columns corresponding to number of Participating Morphemes (PM), Number of Non-Participating Morphemes (NP), and number of Exceptions (Ex). The phonological domain will be selected in each case to maximize this score.

The ratio value measure discussed in Section 3 is computed by dividing the number of participating morphemes within the phonological domain by the total number of morphemes in that domain (participating morphemes + exceptions). As an example, setting this value at an arbitrary 65% threshold gives the patterns in Table 10, and the following segment distribution:

(12) t (3); n (1); k/g (3); r (1); j (2); ʔ (3)

[9] Certain languages in the sample retain a degree of ambiguity even with respect to the proposed quantitative analysis. Take the case of Plains Cree. Prefixes in both nominal and verbal paradigms involve /t/ epenthesis, and number 14 in total, well in excess of the 4 morpheme threshold of the diagnostic which gives the distribution in (14). It turns out, however, that all the prefixes belong to the class of possessive pronoun, that the verbal prefixes are largely homophonous with the nominal prefixes, and half the forms within a single paradigm involve circumfixation, where the prefixal component is identical with another bare prefix form in the paradigm (see Appendix). This calls into question the uniqueness of these morphemes, and whether they can count equally towards satisfying the productivity requirement. One might, conservatively, put the count of unique participating morphemes at 3, not 14. The situation is similar for Waropen, where the similarity of the class of person pronouns raises the possibility of a deletion rather than an epenthesis analysis. Finally, the /n/ epenthesis pattern in Dutch as described in Booij (1995) occurs only for schwa vowel contexts. Furthermore, it is only one option among two other possible repairs: epenthesis of /ʔ/ across word boundaries (when enclitization fails to happen), or schwa deletion.

Table 10
PM/(PM+Ex) ≥ .65

	Seg.	Language	Phonological Domain	PM	NP	Ex
35	k	Dakota	3rd personal plural	1	5	0
36	t	Maori	Trimoraic roots of light syllables in the Passive & Gerund*	2	7	0
37	r	S. Tati		1	3	
38	t	Cree	personal possessive prefixes	14	14	2
39	n	Dutch	ə-final host+clitic†	17	3	0
40	k	Waropen	verbal person prefixes	7	5	2
41	g	Buryat	verbal suffixes on long-vowel final stems	5	7	1
42	t	A. Campa	Verbal Suffixes	14	11	5
45	j	SE Armenian	1-syllable words	4	6	0
50	ʔ	Selayarese	V+V;V=V	5	1	0
54	j	Berber	Suffixation	7	2	2
55	ʔ	Tamil	compounds	1	0	0
56	ʔ	Malay	prefixation	2	0	0

What can be immediately observed for this example threshold is that the number of qualifying languages is quite small: 13 patterns out of an original possible 56. This small number makes generalizations difficult. It can be seen, however, that /t/ epenthesis cannot be said to dominate. There are equal numbers of /t/ and /k/ epenthesis in (12). This outcome, therefore, cannot be taken as support of the hypothesized preference for the coronal place of articulation over the dorsal (Kean 1975, Paradis & Prunet 1991, de Lacy 2006) within certain generative theories of markedness[10]. Of course, we have not

* Adopting the analysis of de Lacy (2003)
† Optional variant of three possible repairs
[10] This is not to suggest that there might not be evidence for a /t/ preference in other phonological domains, or that markedness plays no role in grammar, only that the markedness scale in which /t/ ⟩ /k/ cannot be argued for on the basis of the epenthesis data presented here.

addressed the possibility that there might exist some diagnostic, so far undiscovered, that will demonstrate such a preference.

Let us begin with the premise that Axininca Campa is a strong case of /t/ epenthesis. Setting a diagnostic measure to ensure this outcome will have certain unavoidable outcomes if applied consistently. In particular, such a diagnostic is incapable of excluding all velar epenthesis languages.

To show this, let us first attempt to exclude the Buryat pattern from contention. One way we might do this is to make a strict requirement regarding an invariable phonetic realization of the epenthetic segment[11]. Keeping only the most basic prosodic requirements (which will continue to rule out the languages in Table 6) but removing all other restrictions, the former move will result in a typology of all the languages in Table 9 minus Buryat (and Greek, where the segment is contextually realized as [j]), reflecting the following distribution

(13) t (3), k (2), r (1), n (1), w (1), j (4), ʔ (8)

As shown in (13), this restriction alone does not eliminate all cases of /k/ epenthesis, nor does /t/ emerge as significantly more likely than other segments.

[11] As mentioned above, the epenthetic segment in Buryat varies in place and manner according to its context: [g] before front vowels; the uvular fricative [ʁ] between back vowels, and the uvular stop [ɢ] after front and before back vowels. I will argue that this process is best described as phonetic, rather than phonological, and does not disqualify Buryat from inclusion in Table 9 under C5. This decision is based on the following considerations. One, the pattern of assimilation is quite distinct from that seen in the languages of Table 8; the varying phonetic surface forms are transparently related to an underlying place of articulation that, at the very least, is further back than coronal. Furthermore, phonetic assimilation is likely to be more pronounced for dorsal than coronal segments. Many studies find a large coarticulatory influence on dorsals from following vowels (e.g., Fowler 1984). This is also reflected in the asymmetrical misperception rates for dorsals versus coronals (Winitz et al. 1971, Plauche et al 1997, Guion 1998). Thus, an asymmetry between the dorsal and coronal places of articulation might be expected on phonetic grounds alone, as distinct from the universal phonological principles that are of interest here. Finally, descriptions of phonetic processes are often missing from standard grammars such that our confidence that other languages in Table 9 do not also show similar levels of assimilation should not be overly high.

The other strategy that could eliminate Buryat – disallowing phonological domains defined by vowel length –, aside from being rather arbitrary, once again leaves behind Dakota and Waropen as /k/ epenthesis candidates. Furthermore, it seems to suggest that any kind of conditioning is out. If so, then epenthesis becomes almost vanishingly rare within the sample (Misantla Totonac: ʔ).

Finally, with the stated desire of eliminating all potential /k/ epenthesis candidates, the absolute threshold of productive morphemes could be deliberately set at 8. This leaves only the languages in (14). It can be seen that setting the threshold to eliminate /k/ epenthesis affects the distribution across the board. While it is true that /k/ epenthesis languages have now been eliminated, the total numbers are considerably smaller than before, again precluding generalizations. The situation only worsens if we conservatively remove the somewhat ambiguous cases of Cree and Dutch (see footnote 9), leaving only one instance each of /t/, /j/ and /ʔ/ epenthesis.

(14) Cree (t), Dutch (n), Axininca Campa (t), M. Totonac (ʔ), and Turkish (j)

6. Analysis

It is clear that different diagnostics can produce different typological conclusions from the same sample. However, certain outcomes are consistent. This set of data suggests that robust "true" epenthesis might be much rarer than has been suggested, especially of non-approximant segments. And, secondly, a preference for /t/ over /k/ as a default epenthetic segment must at least be called into question.

Recall that not all languages from Table 5 were investigated, and that a number of other epenthetic segments have been proposed (l, x, ʃ, etc.). The low rate of verification

suggests that the majority of such languages would fail to meet the criteria in (7); however, one or two examples from this set might emerge. This fact, coupled with the very low incidence rate, leaves open the possibility of a fairly uniform distribution over possible epenthetic segments. Even in the absence of any productivity requirements at all, that is, allowing all the cases in Table 9, the distribution may be seen to skew towards ʔ to a certain degree, (15), but other possible differences are obscured by the sampling bias towards /k/, and even more so, /t/.

(15) t (3), k/g (3), ɣ (1), r (1), n (1), w (1), j (4), ʔ (8)

While a markedness preference for the glottal stop is a possibility (as argued for in Steriade 2001, Lombardi 2002, de Lacy 2006 and others), a preference for /t/ otherwise, much less a full scale of place of articulation preferences is not supported by this sample. In turn, this argues against a set of un-rerankable context-free markedness constraints – or any theoretical device that enforces a universal preference for coronal over dorsal places of articulation in the domain of epenthesis.

6a Diachrony

A general '/t/ preference' can be retained however, by situating it upstream from the grammar itself. Rather than a categorical ban on certain grammars, the markedness hierarchy re-surfaces as a bias towards a particular set of grammars, a bias that is always active, but not always realized in the outputs of a particular synchronic grammar (cf. Wilson 2006). Thus, potential filters on possible grammars are 'bottom-up', applying during the process of transmission and learning, not 'top-down' from universal and innate constraints.

The burden of explanation is then shifted to a diachronic theory of the emergence of the given typological distribution. The fact that all non-glide segments may be equally likely is feasible if those segments arose via a historic process of consonant deletion (Blevins 2008). Yet a bias could emerge if certain segments are more likely, for phonetic reasons, to delete, or be 'masked' by neighboring articulations (see Jun 2004). Additionally, the hypothesis of an alternative diachronic source for glide epenthesis could go some way to explaining the differences in distribution observed in the sample. If coarticulation between neighboring vowels were responsible for epenthetic glides (see Browman & Goldstein 1990, Gick 2003), rather than deletion, then the "assimilative" nature of such systems is expected. The context-conditioned examples of ʔ epenthesis could be explained by the same mechanism; in this instance re-timing between adjacent vowels to *avoid* coarticulation would produce the percept of the glottal stop.

The diachronic model also allows for the possibility of three general types of outcome in the non-glide case: a synchronic deletion analysis (no change), a synchronic epenthesis account (requiring re-analysis of a historic deletion process), and something in between. The intermediate possibility is attractive for those languages which fail to meet a subset of the criteria in (7), yet do seem to exhibit some of the traits of epenthesis. Even while the best synchronic analysis of such systems remains unclear, the diachronic approach offers an explanation for their existence, as well as the form of observed alternations.

Finally, the diachronic account for the non-glide epenthesis outcome requires what is known in the historical literature as a 'rule inversion' where deletion in context A becomes epenthesis in context ¬A (see Vennemann 1972). In OT terms this would

require a re-ranking of MAX and DEP. If such a large restructuring of the grammar were difficult to achieve (i.e., required a large burden of evidence) then such paradigms might be expected to be rare – consistent with the failure of most languages of the sample to meet productivity criteria. This is ultimately a learning question (see Section 3a), and further discussion is beyond the scope of the current paper, but see Morley (submitted) for an analysis of exactly this trajectory as a promising explanation for the typological facts presented here.

6b Synchrony

One claim of this paper is that un-rerankable context-free markedness constraints are inadequate to explain the typology of consonant epenthesis. The outlines of a diachronic model of the origins of epenthesis patterns was sketched above. This section details possible synchronic grammars that could result from that learning process. In the first instance, the process of epenthesis (or deletion) will need to be tied to specific contexts. This can be done with a set of MAX/FAITH constraints, indexed to particular morphemes, morpheme positions, and/or morpheme classes (see, e.g., McCarthy & Prince 1995, Casali 1997).

As for the identity of the epenthetic consonant itself, since it is unpredictable, it cannot be computed by the grammar, but must be supplied by the lexicon. Here, the original FILL constraint of Prince & Smolensky (1993/2004) provides a congenial notation, specifing only whether or not a slot is filled with underlying segmental material. I tentatively propose the following family of positive constraints based on this formulation: "$\square_C$/X": fill an empty consonantal slot with X.

For Axininca Campa the split behavior between epenthesis in suffixing environments and deletion in prefixing environments can be captured with the series MAX(Root) >> MAX(MI) >> MAX. The first constraint prevents deletion of any part of the verbal root as a repair for an onsetless syllable. The second constraint prevents deletion of the first vowel (Morpheme Initial) of the suffix (Casali 1997). See Table 11. Nasals agree with the place of articulation of following stops.

Table 11. OT Analysis of Axininca Campa epenthesis pattern

/in + koma + i/	AGREE(Nasal-Place)	□$_C$/t	MAX(Root)	ONSET	MAX(MI)	DEP	MAX
/iŋkomi/			*!		*		*
☞/iŋkomati/						*	
/iŋkomai/				*!			
/iŋkoma/					*!		*
/iŋkomaki/		*!				*	

This ranking of the MAX constraints also prevents deletion of any but the prefix-final vowel under hiatus across the morphological boundary. See Table 12.

Table 12. OT analysis of Axininca Campa deletion pattern

/no + ir + i/	AGREE(Nasal-Place)	□$_C$/t	MAX(Root)	ONSET	MAX(MI)	DEP	MAX
☞/niri/							*
/notiri/						*!	
/noiri/				*!			
/nori/			*!		*		*
/nokiri/		*!				*	

Buryat shows the activity of a similar family of faithfulness constraints. Instead of requiring /t/ in epenthesis situations, however, a velar segment is epenthesized. See Table 13. Long vowels cannot be deleted (shortening does not improve the output), hence the undominated status of MAX(V:).

Table 13. OT analysis of Buryat epenthesis pattern

/yglø: + y:r/	MAX(V:)	□$_C$/velar	ONSET	DEP	MAX
/ygly:r/	*!		*		*
☞/yglø:gy:r/			*	*	
/yglø:y:r/			**!		
/yglø:r/	*!		*		*
/yglø:ty:r/		*!	*	*	

For root-final short vowels, deletion occurs instead. Since there are no short-vowel initial suffixes that attach to short-vowel final stems, V_1 deletion can follow automatically from MAX(V:). See Table 14.

Table 14. OT analysis of Buryat deletion pattern

/axa + a:r/	MAX(V:)	□$_C$/velar	ONSET	DEP	MAX
☞/axa:r/			*		*
/axaga:r/			*	*!	
/axaa:r/			**!		
/axar/	*!		*		*
/axata:r/		*!	*	*	

7. Conclusions

The first contribution of this work is in drawing attention to the necessity of (quasi) theory-independent, consistently applied diagnostics for evaluating linguistic patterns. This is critical for maintaining the falsifiability of linguist theories. Similarly, in the course of the typological study in this paper, a number of issues general to phonological research have surfaced. In particular, the lack of a formal methodology in many critical stages of analysis, and the high degree of ambiguity inherent to any sample of linguistic data proved to significantly complicate the discussion of epenthesis systems.

At root, synchronic theory, and the analyses that underlie it are developed in large part from typological evidence. Like a hypothetical learner, the linguist analyzes a sample of data (although from multiple languages), and generalizes from that data to

make predictions about possible and impossible forms (in any human language). Reasoning from data that is in some way non-representative could lead to serious flaws in the resulting theory. Although self-evident, this is a fact which often does not receive enough attention in synchronic theory, where there is surprisingly little discussion of the size, constitution, or reliability of the typological sample.

In order to establish the set of facts that are to be explained in the domain of consonant epenthesis, this paper supplied an independent typology, created by examining a set of original source grammars. Although far from comprehensive, at 56 patterns, this typology represents a significant increase over typical sets discussed with respect to questions of universality. Furthermore, the data given here are supplied at a level of detail that is frequently missing from analyses that idealize or simplify language patterns.

The importance of such a careful analysis is readily apparent from the failure of these data to conform to the predicted markedness scale (/t/ ⟆ /k/). At the least, it seems that the case for a universal preference for /t/ epenthesis has been greatly over-stated in certain quarters. And, in fact, the case for epenthesis at all (for a given definition of that term), may have been significantly over-estimated as well. Within the margins of sampling error, and under a range of diagnostics, 'impossible' patterns like /k/ epenthesis are about as likely as the 'preferred' /t/ epenthesis (and perhaps most other non-approximant segments) but neither are particularly likely at all. Under this assessment of the data, the theory is responsible for quite a different set of facts, and as a result, might be expected to be modified accordingly.

In the previous sections some suggestions were put forth for just such modifications. A case was made for a diachronic explanation of the observed typology.

Ultimately, this would be a model that integrates both diachrony and synchrony, providing an explicit and formal characterization of the form of grammars that are produced over time by the transmission of surface forms from speaker to listener. Additionally, some proposals were made for a synchronic instantiation of the results of historic changes. Additional work is currently underway to more fully investigate, via computational methods, the type of sound changes and learners that would be necessary or sufficient to produce a synchronic epenthesis grammar.

Acknowledgments

This work was supported by an NSF IGERT grant and a Department of Education Javits Fellowship as well as the Targeted Investment in Excellence from the OSU College of Humanities. I would like to thank Paul Smolensky for helping make this work come about. Thanks also to Cynthia Clopper, Adam Buchwald, and Jenny Culbertson for their comments and suggestions. Matt Goldrick deserves thanks for being excellent in every way, and Brett Hyde for duty above and beyond in providing critical guidance.

Anceaux, J. C. (1961). The linguistic situation in the islands of Yapen, Kurudu, Nau and Miosnum, New Guinea, S-G Ravenhage - Martinus Nijhoff.

Asher, R. E. and T. C. Kumari (1997). Malayalam, Routledge.

Bauer, W. (1993). Maori. New York., Routledge.

Bhat, D. and M. Ningoma (1997). Manipuri Grammar, Lincom Europa.

Blevins, J. (2008). Consonant epenthesis: natural and unnatural histories. In J. Good (ed. Language Universals and Language Change: Oxford University Press. 79-109.

Booij, G. (1995). The phonology of Dutch, Clarendon Press.

Broselow, E. (1984). Default consonants in Amharic morphology. Papers from the January 1984 MIT Workshop in Morphology: 15-32.

Browman, C. P. and L. M. Goldstein (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston and M. E. Beckman (eds.) Papers in laboratory phonology I: Between the grammar and the physics of speech: Cambridge University Press. 341-376.

Burling, R. (1966). The addition of final stops in the history of Maru (Tibeto-Burman). Language 42(3;1): 581-586.

Christdas, P. (1988). The phonology and morphology of Tamil. Linguistics, Cornell.

Crowhurst, M. J. (1994). Foot extrametricality and template mapping in Cupeno. Natural Language and Linguistic Theory 12(2): 177-201.

de Lacy, P. (2003). Maximal Words and the Maori Passive. MIT Working Papers in Linguistics 44: 20-39.

de Lacy, P. (2006). Markedness: Reduction and Preservation in Phonology, Cambridge University Press.

Ebert, K. (1996). Kodava, Lincom Europa.

Elbert, S. H. and M. K. Pukui (1979). Hawaiian Grammar, University of Hawaii Press.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. Perception and Psychophysics 36(4): 359-368.

Geiger, W. (1995/1938). A Grammar of the Sinhalese Language, Asian Educational Services.

Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. In J. Local, R. Ogden and R. Temple (eds.) Papers in Laboratory Phonology VI: Constraints on Phonetic Interpretation.: Cambridge University Press. 222-236.

Gieser, C. R. (1970). The Morphophonemic system of Guininaang (Kalinga). Philippine Journal of Linguistics 1: 52-68.

Glyn Jones, W. and K. Gade (1981). Danish: A Grammar. Copenhagen, Gyldendalske Boghandel.

Gowda, K. S. G. (1975). Ao grammar, Central Institute of Indian Languages.

Guerssel, M. (1986). Glides in Berber and syllabicity. Linguistic Inquiry 17(1): 1-12.

Guion, S. G. (1998). The role of perception in the sounds change of velar palatalization. Phonetica 55: 18-52.

Hahn, R. F. (1991). Spoken Uyghur. Seattle, WA, University of Washington Press.

Hahn, R. F. (1992). Modern Uyghur y~r insertion. Acta Linguistics Hafniensia 24: 77-96.

Hale, K. (1973). Deep-surface canonical disparities in relation to analysis and change: an Australian example. Diachronic, Areal and Typological Linguistics: 401-458.

Hall, R. A. (1953). Haitian Creole. Memoirs of the American Folklore Society 43.

Hayes, B. and M. Abad (1989). Reduplication and Syllabification in Ilokano. Lingua 77: 331-374.

Hdouch, Y. (2008). Extraprosodicity and syllable structure in Berber, LINCOM Europa.

Held, G. J. (1942). Grammatica van het Waropensch.

Herslund, M. (2002). Danish, LINCOM Europa.

Hill, J. H. (2005). A Grammar of Cupeno, University of California Press.

Jaggar, P. J. (2001). Hausa, John Benjamins.

Jarosz, G. (2006). Richness of the base and probabilistic unsupervised learning in Optimality Theory. Proceedings of the Eighth Meeting of the ACL Special Interest Group in Computational Phonology and Morphology, New York City.

Jilali, S. (1976). A Phonological study of Tamazight Berber: dialect of the Ayt Ndhir. Linguistics, UCLA.

Joseph, B. D. and I. Philippaki-Warburton (1987). Modern Greek, Croom Helm Ltd.

Jun, J. (2004). Place Assimilation. In B. Hayes, R. Kirchner and D. Steriade (eds.) Phonetically Based Phonology: Cambridge University Press. 58-86.

Ka, O. (1994). Wolof Phonology and Morphology, University Press of America.

Kean, M.-L. (1975). The Theory of Markedness in Generative Grammar. Cambridge, MIT.

Kenstowicz, M. and C. Kisseberth (1979). Generative Phonology, Academic Press.

Kisseberth, C. (1970). On the functional unity of phonological rules. Linguistic Inquiry 1: 291-306.

Laidig, C. (1992). Segments, syllables, and stress in Larike. In D. A. Burquest and W. D. Laidig (eds.) Phonological studies in four languages of Maluku. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington Publications in Linguistics. 108. 67-126.

Lee, D.-Y. (1998). Korean Phonology, LINCOM Europa.

Leslau, W. (1997). Ethiopic Documents: Argobba. Wiesbaden, Harrassowitz Verslag.

Lombardi, L. (2002). Coronal epenthesis and markedness. Phonology 19: 219-251.

Luthy, M. J. (1973). Phonological and lexical aspects of colloquial Finnish. Bloomington, Indiana University.

MacKay, C. T. (1994). A Sketch of Misantla Totonac Phonology. International Journal of American Linguistics 60(4): 369-419.

MacKay, C. J. (1999). A Grammar of Misantla Totonac. Salt Lake City, The University of Utah Press.

Malikouti-Drachman, A. (2009). Cross dialect ambiguities of epenthesis. Journal of Greek Linguistics 9.

McCarthy, J. J. and A. Prince (1994). The emergence of the unmarked: optimality in prosodic morphology. Proceedings of the North East Linguistic Society.

Mielke, J. (2007). P-Base.

Mithun, M. and H. Basri (1986). The phonology of Selayarese. Oceanic Linguistics 25(12): 210-254.

Natarajan, G. V. (1985). Abujhmaria Grammar. Mysore, Central Institute of Indian Languages.

Newton, B. (1972). The generative interpretation of dialect: a study of modern Greek phonology, Cambridge University Press.

Onn, F. M. (1976). Aspects of Malay Phonology and Morphology. Linguistics. Urbana, ILL, University of Illinois at Urbana-Champaign.

Paradis, C. and J.-F. Prunet (1991). Asymmetry and Visibility in Consonant Articulations. In C. Paradis and J.-F. Prunet (eds.) The Special Status of Coronals: Internal and External Evidence: Academic Press, Inc. 2.

Parker, S. (1989). The sonority grid in Chamicuro phonology. Linguistic Analysis 19: 3-58.

Paster, M. E. (2006). Phonological Conditions on Affixation. Linguistics, Universaity of California, Berkeley.

Payne, D. L. (1981). The Phonology and Morphology of Axininca Campa, University of Texas at Arlington.

Piggott, G. L. (1980). Aspects of Odawa Morphophonemics, Garland Publishing.

Plauche, W., C. Delogu, et al. (1997). Asymmetries in consonant confusion. Eurospeech 97, Rhodes, Greece.

Poppe, N. N. (1960). Buriat Grammar, Indiana University Publications.

Prince, A. and P. Smolenksy (2004). Optimality Theory, Blackwell Publishing.

Radhakrishnan, R. (1981). The Nancowry word: phonology, affixal morphology and roots of a Nicobarese language, Linguistic Research Inc.

Renaud, K. Y.-K. (1974). Korean Consonantal Phonology. Linguistics, University of Hawaii.

Rischel, J. (1974). Topics in West Greenlandic Phonology. Copenhagen, Akademisk Forlag.

Shaw, P. A. (1980). Theoretical Issues in Dakota Phonology and Morphology, Garland Publishing, Inc.

Shetler, J. (1976). Notes on Balangao Grammar. Huntington Beach, CA, Summer Institute of Linguistics.

Spring, C. (1990). Implications of Axininca Campa for prosodic morphology and reduplication. Linguistics, University of Arizona.

Steriade, D. (2001). The phonology of perceptibility effects: the P-map and its consequences for constraint organization. UCLA.

Stonham, J. (1999). Aspects of Tsishaath Nootka phonetics and phonology, LINCOM Europa.

Sulkaka, H. and M. Karjalainen (1992). Finnish. London, Routledge.

Tesar, B. and P. Smolenksy (1998). Learnability in Optimality Theory. Linguistic Inquiry 29(2): 229-268.

Tranel, B. (1981). Concreteness in Generative Phonology: Evidence from French, University of California Press.

Underhill, R. (1976). Turkish Grammar, The MIT Press.

Vaux, B. (2001). Consonant epenthesis and hypercorrection. unpublished.

Voegelin, C. F. (1935). Tubatulabal Grammar. University of California Publications in American Archeology and Ethnology 34(2): 55-190.

Wali, K. (2005). Marathi, Lincom Europa.

Watson, J. C. E. (2002). The Phonology and Morphology of Arabic, Oxford University Press.

Winitz, H., M. E. Scheib, et al. (1971). Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech. Journal of the Acoustical Society of America 51(4(2)): 1309-1317.

Wolfart, H. C. (1973). Plains Cree: a grammatical study. Transactions of the American Philosophical Society 63(Part 5).

Woollams, G. (1996). A Grammar of Karo Batak, Sumatra. Canberra, Pacific Linguistics.

Yallop, C. (1977). Alyawarra, an Aboriginal language of central Australia. Canberra, Australian Institute of Aboriginal Studies.

Yar-Shater, E. (1969). A grammar of Souther Tati Dialects. The Hague, Mouton.

Zoll, C. (1998). Parsing below the segment in a constraint-based framework. Stanford, CA, CSLI Publications.