Can phonotactic constraints inhibit segmental change? Arguments from lenition and syncope.

Patrick Honeybone
University of Edinburgh, patrick.honeybone@ed.ac.uk

Abstract
This article considers the interaction of phonotactics and diachrony. I argue two things: (i) language-specific phonotactic constraints on phonological forms can inhibit otherwise regular innovations, and (ii) the fact that such phonotactically-motivated process inhibition occurs in historical phonology is itself evidence for the reality of phonotactic constraints. I assume that there is a difference between those gaps in a language's lexicon which are due to chance ('A-gaps') and those which are ruled out by the grammar ('S-gaps') and I consider some evidence in favour of this view. I consider two case studies where an understanding of phonotactics is necessary to analyse the patterning of change: Mid-Scots θ-debuccalisation and a late Middle English syncope. I ground the discussion in arguments about what phonotactic constraints are, and how they can be involved in diachrony. This involves a consideration of a number of examples from English, including onset-OCP-related constraints, the OCP(sibilance) constraint, and the constraint which imposes the defective distribution of [h].

Keywords: phonotactics; historical phonology; lenition; syncope

1. Introduction

What is the status of sequences of segments that do not occur in the phonology of a language?[1] One common position is that at least some such missing sequences (or 'gaps') are forbidden by the *phonotactics* of languages' phonologies. If this is right, then the phonotactic entities that enforce such gaps must form part of a speaker's phonological knowledge, and must behave like other aspects of phonology – for example in the interaction with acquisition and diachrony. It is this latter point that I focus on in this article. I consider what can occur in the diachrony of phonotactics, and how they might interact with other aspects of phonology in cases of phonological change. If phonotactic entities *are* as psychologically real as other kinds of phonological object (such as segments, stress and feet), then we should expect that their effects *will* be detectable in cases of phonological change; conversely, if it can be shown that phonotactic entities play a role in conditioning phonological change, and have a definite diachrony of their own, that is in itself evidence for their phonological reality.

The article is structured thus: Section 2 considers some of the fundamental notions behind phonotactics and explores some examples from English, setting out the groundwork; Section 3 brings diachrony into the picture; and Section 4 focuses on my main question: can phonotactic constraints inhibit segmental change? (I think that they can.) Section 5 concludes. The empirical base for the article comes mainly from English (understood broadly to include all 'insular' developments of West Germanic), but the ideas should be relevant to all languages. In Section 4, I focus on the analysis of two case studies where an understanding of phonotactics is necessary in order to understand the patterning of change: 'Mid-Scots θ-debuccalisation' and pre-z 'Spätmittelenglischer Schwund' (a case of vowel reduction/syncope).

[1] A version of this paper was presented at the Diachronic Phonotactics Workshop at the University of Vienna in September 2017. I am very grateful to the workshop's organisers and its other participants for the chance to reflect on the issues discussed here, and for the comments and questions that I received there. The paper has grown in conception since then and I am grateful for the comments of two anonymous reviewers, and especially to Niki Ritt for his detailed and insightful consideration of the piece and for many suggestions for clarification. All of this has considerably improved the piece, and I am sorry that it still perhaps tries to say too much. I alone am responsible for this (and also, of course, for all the arguments that I put forth in it).

2. What is/are phonotactics?

If we conceive of phonology as a grammatical system which determines which structures are allowed and which are not, then the question of 'what *cannot* occur' is a crucial one for understanding phonotactics. With a perspective like this, the question asked at the start of this article is a central one: are sequences of segments absent from languages because they are forbidden by the phonological grammar? A common answer is that some are and some are not, giving two fundamental types of gaps (of absent sequences in a language) – some are 'accidental' gaps because there is nothing in the phonology of a language which forbids them but there happens not to be a word or morpheme in the language that attests the sequence. Others are gaps because no word or morpheme in the language is *allowed* to feature them, because the grammar forbids it – these are called a range of things in the literature: 'systemic gaps' or 'systematic gaps' or 'structural gaps'. Because all of these start with an ‹s›, I will call them 'S-gaps', and because of that, I will call the former 'A-gaps'. So: are sequences of segments absent from languages because they are forbidden by the phonological grammar? Only S-gaps are forbidden by the grammar, and the relevant part of our knowledge of phonology can thus assumed to be *phonotactic knowledge*.

The assumption that there are indeed S-gaps opens up a range of types of evidence that might demonstrate the difference between the two types of gap. This cannot all be considered here, but one kind of evidence that will feature below involves loanwords: if a non-occurring sequence is forbidden by the phonological grammar, then we would expect that new words that enter the language will not be allowed to feature the sequence – if a loanword does feature a sequence that would be ungrammatical (because it comes of a language with a different grammar), the prediction is that it will be adapted by the borrowing language to fit with its S-gaps. If a loanword features an absent sequence that fills an A-gap, on the other hand, it will be adopted into the borrowing language unchanged. The evidence from loanword adaptation requires careful interpretation – like everything, it is not as cut-and-dried as I set it out here (see, for example, Peperkamp 2005), but a substantial strand of literature argues that there is good reason to recognise that phonological principles play a key role in determining what happens to loanwords (see, for example, Paradis 2006). The fact that words which feature sequences that are not found in a borrowing language (of the S-gap type) *are* adapted when they are borrowed, while words featuring other non-occurring sequences (of the A-gap type) are *not* adapted, is important evidence to show that S-gaps are psychologically real, because it shows that they are phonologically productive. The case studies of particular phonotactics that I consider below (especially in Section 2.2) will offer some evidence in favour of the idea that S-gaps are productively enforced in loanword adaptation, while A-gaps can freely be filled in loanword adoption.

The 'grammar-based' approach, which assumes that phonotactics is fundamentally about establishing which S-gaps languages have (and about generalising over this to understand how such S-gaps work in phonology) makes further investigable predictions in terms of the phonological behaviour of S-gaps: for example, we should expect their signature to be visible in the patterning of phonological change, and that they have a diachrony of their own. This is the central point in this paper (from Section 3 onwards). My title asks 'Can phonotactic constraints [which we can now also call 'S-gap-enforcing constraints'] inhibit segmental change?', and I will conclude by arguing that there is evidence that they can, and that such evidence shows that we do indeed need the concept of the S-gap. The implication of this is that the investigation of diachrony can provide important evidence for our understanding of phonology in general.

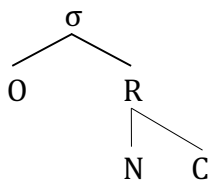
A specific grammatically-enforced gap in a language is sometimes (perhaps confusingly) described as 'a phonotactic' (shorthand for 'a phonotactic generalisation') so we can ask both *what phonotactics is* (as in this current section) and what the phonotactics are in any language (as in Section 3). In what follows, I assume that S-gap-enforcing phonotactics is/are implemented by psychologically real phonological entities. The paper's title assumes that they

exist in the form of constraints which forbid specific types of sequences – generalisations about the grammatical (and ungrammatical) distribution of segments in a language which are implemented by 'static' phonological entities (ones which are not 'dynamic' in the sense that they do not describe how distributions are brought about by processes). This is the default assumption in current formal phonological theory, of both constraint-based and rule-based types, as I consider in the following section.

2.1. Modelling phonotactics phonologically

This paper implements phonotactics as negative grammatical constraints which limit the phonological form that words (or other elements such as morphemes) can take. The slots provided by basic syllable structure have shown themselves to be so useful in describing phonotactic generalisations that most phonotactic work uses them without comment, or ends up reinventing them by talking about 'word/morpheme initials' instead of 'onsets' or 'word/morpheme finals' instead of codas or rimes.[2] Basic syllabic constituents will prove indispensable below, and I argue in favour of using them where it becomes appropriate. This means that 'word-like' gaps, such as [bnɪk] are not S-gaps in their entirety – the rime is fine and the problem is with the onset (which could therefore lead a speaker to reject the whole string if asked about it). A reasonable hypothesis is that phonotactics are (co-)occurrence constraints that refer to specific syllabic constituents, and I adopt this position below. I assume that basic syllable structure is a given, providing structural slots into which segments may be fitted. I follow one traditional syllabic model, which assumes that syllables branch to form constituents, that the syllable is composed of an onset and rime, and that the rime may branch to form a nucleus and a coda (and that the nucleus is the head of the rime), as shown in (1). This is all contentious in theories of syllabic/skeletal structure, but little of what follows relies on the precise nature of these assumptions, and the reader will be able to translate these notions into their own theory if it differs.

(1)



I assume that some things are basic structural facts in English: two-member onsets and coda-consonants are possible (and this is interesting because other languages are phonologically different in this respect: some only have one-member onsets, and others do not have codas). These 'basic structural facts' are themselves imposed by general constraints on syllabic structure (either language universal, in requiring a nucleus, for example, or English-specific in allowing codas). With this in place, the phonotactic question is: do all segments occur in all of the structural slots that the language allows? I focus below on the occurrence of consonants.

Putting everything together, we could represent phonotactic constraints as in (2), which accounts for one of the observations just made (that [bnɪk] in not possible in English). This adopts the practice made common in Optimality Theory of writing constraints using small capitals.

[2] I follow the convention of using the spelling 'rime' for the syllabic constituent, to differentiate it from the poetic notion of 'rhyme', which is not the same thing.

(2) *ONSET-bn
the sequence [bn] is forbidden in an onset

Whether we would want to represent the observations just made exactly as in (2) is another matter. While not literally incorrect, the constraint in (2) misses the point – we should want constraints to explain phonotactic prohibitions in their most general and incisive form. We will not adopt the precise constraint in (2) in what follows, but we will adopt the formalism used to express it.

A crucial final point in any general consideration of phonotactics is: where do they apply? Early generative work on the issue was clear that Morpheme Structure conditions constrained Underlying Representations – how forms can be stored. In OT, phonotactic constraints, like all constraints, evaluate surface forms. Which is right? (Is only one right?) Shibatani (1973) offers a classic argument in favour of 'Surface Phonetic Constraints' (that is, that there must be constraints that hold at the surface level of representation), pointing out that, in a language like German, which has Final Obstruent Fortisisation (*Auslautverhärtung*), such that no lenis obstruents may occur at the right-edge of a word in surface forms (but can occur there in underlying forms, as in /bʊnd/ → [bʊnt]), speakers have intuitions that forms like [bʊnd] and [liːb] are non-German.[3] Furthermore, loanwords from English (where final lenis consonants are possible) are adapted into German using final fortis consonants (as in *Klub*, from *club*, with final [p] and *Jazz*, with final [s]). Both pieces of evidence point to the effect of constraints on surface forms (that is, at the 'end' of phonology) because the gap in the distribution occurs at the surface, not the underlying level. This does not absolutely rule out that there may *also* be constraints on underlying forms (see Faust, Jatteau and Scheer 2018 for arguments that they are indeed necessary), but given that at least some constraints must hold at the surface I model the constraints that I investigate below in that way. More broadly, while I model phonotactics as static constraints on surface forms, I formulate dynamic phonological generalisations ('processes'), where they crop up, as traditional phonological rules, because rules have a wide currency and will express the relevant phenomena clearly. I return to this in Section 2.3, where a full model for these aspects of phonology is given.

2.2. Recognising and modelling English phonotactics

Armed with all this, what *are* the phonotactic constraints of English?[4] There is no shortage of previous work which has insightful things to say about English phonotactics: Kruisinga (1943), Fudge (1969), Hammond (1999), Szigetvári (2007) and Bauer (2015) are some examples. While earlier work (such as Kruisinga 1943) had to rely on manual searches of the English lexicon, it is now not difficult to find non-occurring sequences in English: resources such as the *MRC Psycholinguistic Database* (see Coltheart 1981) and *CUBE* (Lindsey and Szigetvári 2013) offer easily searchable databases that can show in seconds whether any specific sequence is attested in any word of their dictionaries, and the online *Oxford English Dictionary* also provides useful searchable information. I make use of such sources below. When allied with information

[3] This claim about German speakers' intuitions is made by Shibatani without citing a source. Niki Ritt (p.c.) writes that German-speakers can produce final lenis obstruents when they make an effort, such as in primary school when teachers are giving dictation, but that they sound "unnaturally hyper-correct." I assume that the pronunciation described here (which is intended to help pupils spell words correctly in dictation) is similar to that described for 'peripheral' pronunciations of *Knesset* and *raison d'être* in footnote 9.

[4] This naturally varies somewhat from variety to variety, and where there are differences I typically consider a General British-type variety. I only consider a few constraints in this article – it is by no means intended to offer an exhaustive list.

from introspective (conscious) wordlikeness judgements [5] and (subconscious) loanword adaptation, it is possible to sort the A-gaps form the S-gaps. I consider all of the above in this section, and focus on four phonotactic constraints of English in some detail, several of which will be important in Sections 3 and 4, where I consider their interaction with diachrony.

Why or how a learner might assume that a gap is due to a phonological constraint? Lots of sequences do not occur, but it is unlikely that languages have constraints that forbid them all. For a gap to be phonologically interesting, it needs to be plausible (that is, it must be plausible that we might expect to find words with the structure in question because the structure fits in with the 'basic structural facts' of the language) and systematic. No word of English ends with [plksbf], but it would be ill-advised to assume that this is ruled out by a constraint like *CODA-plksbf. The S-gaps of the type considered here are different. They are cross-linguistically possible and are phonologically plausible in English. Because the gaps are phonologically plausible (for example, they are systematic in involving natural phonological classes), it is reasonable to assume that they might be noticeable by learners.

The first example phonotactic that I consider deals with an observation that has often been made about the distribution of English consonants, as in (3), which uses the wording from Hammond (1999: 58).

(3) "... all consonants of English except [h] can appear as a single-consonant coda

[h] proscription
[h] cannot occur in codas.

This is expressable as a *CODA constraint: *CODA/h."

The constraint in (3) passes the test of noticeability. If we assume that consonants at the ends of words are in codas (as most models do), what Hammond says is true (for rhotic varieties, such as the American one that Hammond is describing) – all other consonants occur in words following a vowel at the end of a word, but no single word of English features [h] in that environment. This includes even the most similar consonants to [h], which (assuming that [h] is a fortis fricative) include [f] (as in *laugh*) and [θ] (as in *bath*). Even in rhotic varieties there is only one other consonant missing in this environment ([r], at the surface at least), so it is not unreasonable to assume that a learner will notice the absence of [h]. There is no general principle why [h] should not occur in that environment – other consonants do, and we know that [h] can occur there in other languages (as in Persian: [noh] 'nine', [ʃɒ:h] 'king'). The rational conclusion for a learner is that English does indeed feature a language-specific constraint which constrains the occurrence of [h] – the absence of [h] is a S-gap, not an A-gap. [h] does not occur in a coda following any vowel, so it is systematic.

It furthermore seems right to use syllable structure in the constraint – if we assumed that the constraint were *WORDEND-h, we would predict that [h] should be able to appear in word-internal codas, so sequences like [ah.tə], [ɛh.nɪk] should be possible words (modelled on attested [af.tə] *after* and [ɛθ.nɪk] *ethnic*), but no such sequence is a word in English. I shall refer to the constraint in question as *CODA-h.[6] If we assume that *CODA-h is a psychologically real

[5] A considerable literature exists investigating wordlikeness judgements. I lack the space to consider it here, but rely on such work as Gorman (2013) and Lentz and Kager (2015), who argue that categorical phonotactics play a clear role in them (along with probabilistic knowledge, which is of a different nature).

[6] In fact, the restriction on the distribution of [h] in English is more complex than this. As well as being prohibited in codas, [h] does not occur in certain types of onset, either. Harris (e.g., 1994, 2000) stresses the importance of forms such as the derivationally-linked pair *prohibit*~*prohibition*, which imply that [h] does not occur in unstressed onsets, and invokes foot-structure to account for this (*prohibit* = σ.hσ́.σ, *prohibition* = σ̀.~~h~~σ.σ́.σ). Davis and Cho (2003) point out that the occurrence of [h] in words like *Tarahumara* (σ̀.σ.hσ.σ́.σ) show that things are even more complex, and argue for complex foot structure to account for it. It may well be that there is a constraint banning foot-medial-[h] in English, as well as one banning coda-[h], but the full details of this cannot be explored here. A constraint banning coda [h] is unambiguously needed, as many final consonants are not foot-medial.

part of the phonology of English, we predict that a sequence ending with an [h] will 'sound wrong' in English and that if any such sequence were to be introduced into English as a loanword, that it will be adapted to remove the final [h], as discussed at the start of Section 2 – if *CODA-h really is a synchronically real aspect of our phonotactic knowledge, it should be productive. This is indeed what occurs: [ʃɒːh] 'king' has been borrowed from Persian as [ʃɑː] *Shah* 'a king of Persia or Iran'.

My second and third example phonotactics involve the structure of English onsets. Various phonological classes can be recognized among the consonants of English (for example, the stops, the labials, the obstruents) and an obvious question for our purposes is: to what extent can these classes freely combine? If the occurrence of [s] is set aside (because it is subject to a distinct set of generalisations, some of which also apply to the other fortis sibilant [ʃ], so that is also set aside), English only ever allows two-member onsets, all of which fully follow the Sonority Sequencing Principle. This general principle rules out a vast number of possible onsets (e.g., [lt], [nt], [wt]), but does not account for all the gaps in combinability. One way of setting out the attested two-member obstruent-containing non-sibilant onset sequences of reference varieties of British English is given in (4). This is based on the clusters described (after detailed searching) by Szigetvári (2007) as attested in word-initial position in English, but is set out to make the phonological classes clear.[7]

(4)	pl	bl	fl				kl	gl
	pr	br	fr	θr	tr	dr	kr	gr
	pj	bj	fj	θj	tj	dj	kj	
				θw	tw	dw	kw	gw

An initial observation from (4) is that, once we allow for general gaps (the Sonority Sequencing Principle, the special status of sibilants, and some type of minimal sonority difference, so that only obstruents and non-nasal sonorants can combine), most combinations occur, so that it is plausible that the few which do not occur might be noticeable by learners. Some of the gaps in (4) may well be A-gaps, however. The case of [gj] is instructive. Szigetvári (2007) in fact includes it as a possible sequence of English on the basis of the one word *gules* 'red, as one of the heraldic colours'. The *OED* puts the word in one of the 'less frequent' frequency bands. The *OED* shows that it was borrowed from Old French, but it has been robustly attested in English since the 15th century so we might expect it to have been adapted by now if it conflicts with English phonology. Another way of tackling the issue is to note that a constraint of the type *ONSET-gj would be phonologically very specific, not involving any phonological class – lenis stops can combine with [j], as shown by [bj] and [dj]; velars can combine with [j], as shown by [kj]; and [g] can combine with glides, as shown by [gw] – so a learner might think it unlikely that English would feature *ONSET-gj. All this implies that, for those speakers who do not have *gules* in their lexicon (and there are likely many), [gj] would be an A-gap and we would predict that if such a speaker came across the city of *Gyandzha* in Azerbaijan (see Everett-Heath 2014), they would have no problem in attempting a pronunciation with initial [gj]. As considered above, if the S-gap vs A-gap distinction is real, A-gaps should not be productively enforced when loanwords are borrowed.

This is quite different from the gaps which involve [pw, bw, fw] and [tl, dl, θl]. These sets both involve phonological classes of segments (the labial obstruents and the coronal obstruents,

[7] The table in (4) is based on the table in Szigetvári's appendix 1, but it removes those clusters which are indicated in his appendix 2 as being included only on the basis of 'unique examples' which are clearly loanwords and may be preserving the phonology of the donor language (see footnote 9). It also ignores a few words which have initial clusters which have lenis fricatives as their first member (such as *vlog*, *zloty*) – these words are all marginal in some sense, but their existence may in fact mean that such clusters should be included in (4). The table also ignores the fact that sequences of two sonorants are possible where the second is [j] (as in *mural*, *new*, *lurid*) – some (such as Kruisinga 1943 and Davis and Hammond 1995) see such facts, along with other evidence, as indicating that these forms actually involve a diphthong (of the type /ɪ̯u/). Reconsidering the decisions on these issues would alter the table in (4), but would not change the basic points made here.

respectively), so they are much more plausible candidates as psychologically real phonological generalisations. I discuss the latter here, which is more robust (neither Szigetvári 2007 nor Bauer 2015 find a single attestation of a word with initial tonic-syllable [tl, dl, θl]). I discuss [tl] as an example of this class of gap, but the same points fundamentally hold for the others.

The absence of [tl] as an initial sequence is not language-universal: there are languages which allow initial [tl] (for example Serbo-Croat has *tlo* 'ground', *tlak* 'pressure'). The English gap is indeed related to the onset environment because there is no general absence of [tl] sequences in English – they occur freely in medial positions, including in morpheme-internal medial positions (in words such as *atlas, butler, cutler*). We can see that a syllable boundary splits the sequences into [t.l] in such cases, meaning that the occurrences of [t] are in codas, because they are glottallable or glottaliseable in varieties that allow such phenomena (to stops in codas). It thus seems fair to conclude that English features a constraint along the lines of *ONSET-tl. However, this is too specific. It is widely assumed that the [tl] facts are due to a general constraint which also affects [dl] and [θl], and that this is due to the fact that [t, d, θ] and [l] share the coronal place of articulation (unlike the attested sequences involving [l]). This can be expressed as in (5), which assumes that the constraint is a specific case of the 'Obligatory Contour Principle', which forbids the occurrence of identical feature specifications on adjacent phonological entities. The segments [r] and [j] are also coronal, and can combine with [t, d, θ], so the precise nature of the OCP violation must involve anterior coronals.

(5) OCP-ONSET(CORONAL)
a sequence of two anterior coronal segments is forbidden in an onset

If we assume that OCP-ONSET(CORONAL) is indeed a constraint of English, then we would predict that borrowings into English with an initial /tl/ will be adapted. This seems to be the case. Names for peoples or places in American languages with such sequences, such as *Tlapanec* tend to be realized by English speakers with initial [təl], using an epenthetic schwa to remove the sequence.[8]

A third example phonotactic can also be seen in evidence from loanword adaptation, which implies that the systematic absence of onset sequences featuring a (non-sibilant) obstruent followed by a nasal is enforced by a phonological constraint. It is unnecessary to demonstrate that words with initial ‹pn› have a 'silent p', but it is also notable that more recent loans, such as the brand name *Knorr* (which has initial [kn] in the original German) is pronounced in English with initial [n], using deletion to remove the sequence.[9] This implies that English has a psychologically real (language-specific) constraint which encompasses *ONSET-kn. The true

[8] Such observations are exemplified in the samples submitted to pronunciation websites, such as https://www.howtopronounce.com/tlapanec/ (checked in May 2018).

[9] A reviewer argues that this observation is invalidated by the fact that *Knesset* (the name of the parliament in Israel) is pronounced [knɛsɛt] in English. I do not think this is the case. It is true that formal pronunciations of the word may have the initial cluster [kn], but this is likely due to the fact that speakers are making an effort to preserve (what they assume to be) the original pronunciation of a word from another language, given that it is clearly marked as 'foreign' due to its referent. Essentially this is the direct importation of a word with one language's phonology into the phonology of another language, like when an English speaker pronounces the phrase *raison d'être* with a [ʁ]. This is confirmed by the fact that speakers also keep adapting the word *Knesset* to fit with the ban on onset-kn (when the impetus to pronounce [kn] to preserve its foreignness fails or is absent). Pronunciations of the words as [nɛsɛt] and [kənɛsɛt] are attested, as in the guide to pronunciation in the Wiktionary entry for the word (https://en.wiktionary.org/wiki/Knesset, checked in December 2018), and as demonstrated in the speaker in the 'Oxford Dictionaries' sample pronunciation here: https://en.oxforddictionaries.com/definition/knesset (checked December 2018) – the transcription given there is [knɛsɛt], but the speaker epenthesises a schwa to break up the [kn] cluster. All this shows that *ONSET-kn is productive in English (but, like other aspects of phonology) can be suppressed if a speaker wants to be faithful to a non-English form. A reviewer rightly points out that discussion of issues like these could also invoke the distinction between a phonological 'core' and a phonological 'periphery', as discussed in Prague School work on phonology and elsewhere. This might place *Knesset* as [knɛsɛt] and *raison* (*d'etre*) as [ʁɛzɔ̃] (or even [ʁɛzɒn]) in the periphery, along with the final lenis obstruents in German dictation (from footnote 3).

constraint involved is broader, however, as it is involved in enforcing the requirement for a minimal sonority difference between the two members of a cluster discussed in connection with (4). The sequences [pn, bn, fn, θn, tn, dn and gn] are also absent in onsets, for example, even though sonority sequencing allows them, so the constraint involved may be best represented as something like that in (6), where 'T' stands for any non-sibilant obstruent, and 'N' stands for a nasal.

(6) *ONSET-TN
a sequence of an obstruent followed by a nasal is forbidden in an onset

The fourth and final example phonotactic in this section focuses on the extent to which the segments *within* a phonological class can be combined. The group is 'the sibilants', which are often identified as behaving as a class (Ladefoged and Maddison 1996 use 'sibilant/non-sibilant' as a fundamental classificational category, for example, and trace its use back to Holder 1669; some examples of recent phonological work using sibilant as a phonological class are Evers, Reetz and Lahiri 1998 and Boersma and Hamann 2008). The precise definition of sibilance is controversial; for present purposes, I take it to include all and only the strident coronals. On this basis, English has six sibilant segments: /s, ʃ, z, ʒ, t͡ʃ, d͡ʒ/, and there seems to be a far-reaching constraint on their combinability.

It is perhaps not surprising that no two of the English sibilants can combine in initial position, given that English respects sonority sequencing in onsets (if we set aside [s] and perhaps [ʃ]), so no two obstruents might be expected to be combinable in an onset (let alone two fricatives). We cannot set aside [s] and [ʃ] here, however, and if we consider the wider behaviour of [s], then the absence of sequences of sibilants is more surprising: affricates do not combine with anything in onsets, but fricatives can: [sf] is a possible onset sequence, as in *sphere, sphinx, sphincter*. It is true that all of these are borrowings into English, but none of them show any current sign of being adapted.[10] In total, the *OED* (online) lists 103 words beginning with ‹sph›, from *sphyraena* to *sphyrelaton* (although some of these are morphologically related), and it also lists two words beginning with ‹sth›, *sthenic* and *sthenia*, transcribed with initial [sθ] (with attestations from the late 18th to the late 19th century), tentatively indicating that English does not forbid initial sequences of s+fricative, although there are no words beginning ‹ss› or ‹ssh›. It is clear, however, that initial sequences of sibilants are linguistically possible, for example in Polish [sʃ] (as in *zszywka* 'staple'), [zʒ] (as in *zrzucić* 'to throw down'), and even [ss] (as in *ssać* 'to suck').

This becomes clearer when we recognize that no sequence of any two sibilants occurs morpheme-medially in English – not even across syllables.[11] A search for all possible two-way permutations of the four fricatives ([sʃ, ʃs, zʒ, ʒz, sz, zs, sʒ, ʒs, zʃ, ʃz, ʃʒ, ʒʃ]) in both the *MRC Database* and *CUBE* finds only words where there is a morpheme boundary between the two sibilants (such as [sʃ] in *misshapen*, [ʃs] in *fish-slice*, [zs] in *transsexual*, and [zʃ] in *newssheet*).[12]

[10] These words have been in English for a long time, so we would expect them to have been adapted by now if they are problematic phonologically: the *OED* (online) gives a robust number of attestations of all three words (with ‹sph›) from the middle of the 16th century.

[11] Much work on phonotactics refers to the morpheme as crucial in the definition of relevant domains. For example, Hammond (1999) explicitly restricts his focus to the phonotactics only of monomorphemic words. It is often noted that the constraints that apply within a morpheme can be more restrictive than those that apply across morpheme boundaries – thus, for example, [θ] occurs after obstruents at the end of English words, but only if they are polymorphemic, such as *breadth* or *twelfth*. The issue is complex and would take us too far afield if I tried to do it justice here. It is only in this current case where we need to refer to morphological structure in the formulation of a constraint, but it is indeed an important aspect of the constraint. To get a full picture, it needs to be recognised that it can be crucial to describe whether a constraint is purely phonotactic or, rather, morphonotactic (see, for example, Dressler and Dziubalska-Kołaczyk 2006).

[12] It is also the case that no sequences of an affricate (with a sibilant portion) followed by a sibilant occurs (which rules out such sequences as [t͡ʃs, t͡ʃz, t͡ʃʃ, t͡ʃʒ, d͡ʒs, d͡ʒz, d͡ʒʃ, d͡ʒʒ]).

There is no general ban in English on tautomorphemic sequences of fricatives, however – for example, [sf] occurs in *asphalt, blaspheme*, [fθ] in *diphtheria*, *ophthalmology*, [sθ] in *aesthetic*, *anesthesia*. It is also the case that no sequence of sibilants occurs in final position in English, not even if a morpheme boundary intervenes. This is even more surprising than the above, because sonority plateaus are widely tolerated in this environment in English words, for example [pt] in *rapt*, [kt] in *strict*, [θs] in *meths*, [fθ] in *twelfth*.

Why do sequences such as [ʃs] and [sʃ] not occur in English in any of these environments? Several of the sibilants of English are common, so it is phonologically surprising that they cannot combine. It is often proposed that the reason for this is that English is affected by a further OCP-type constraint, of the type given in (7). As well as [ʃs] and [sʃ], this will rule out other non-occurring sequences, such as [ʒz, zʒ, t͡ʃs, d͡ʒz].

(7) OCP(SIBILANCE)
a tautosyllabic or tautomorphemic sequence of sibilant segments is forbidden

This section has considered some examples of phonotactic constraints from present-day English, featuring a number of well-established cases of S-gaps, which we can reasonably expect a learner to notice, on the assumption that everything that is not forbidden is allowed[13] and that they involve a systematic absence of a plausible combination of segments.

2.3. Implementing phonotactics phonologically

If such constraints exist in speakers' grammars, how do they fit in with other aspects of phonological knowledge? The kind of phonotactic constraints proposed in the previous section are not tied to only one phonological framework (and I see this as a good thing, indicating that they are more likely to be true). The model that I adopt below when I consider the interaction of phonotactics and diachrony is a 'mixed' rule-and-constraint model, but it is worth observing that the key ideas would also translate into an Optimality Theoretic approach because this shows their widespread applicability. OT works exclusively with constraints, and phonotactics of the type just exemplified can be straightforwardly fitted into OT analyses (indeed, this has formed part of the argumentation that has led many to adopt OT).[14] Hammond (1999: 46) shows this clearly, as in the extract reproduced here (using Hammond's precise conventions) as (8), which ties in with the material given in (3).

[13] This is certainly a controversial assumption, as a reviewer points out, but I hope that the discussion above on the productiveness of S-gaps but not A-gaps (in the adaptation of loanwords, for example) shows that sequences which are not forbidden but do not occur can indeed be allowed if that happen to be encountered when a new word enters a language.

[14] It is true that this means that constraints enforcing S-gaps do not have a clear special status in OT, unlike in the model that I go on to develop next (so phonotactic constraints seem to be formally the same kind of thing in OT as the constraints which enforce the 'basic structural facts'). There may still be a difference between the types of constraints, however, with phonotactic constraints being those which refer to impermissible combinations of specific segments and structural slots.

(8) " ... *Ahab* [ehæb] is unambiguous in its syllabification because of *CODA/h.

Unambiguous syllabification of Ahab

	/ehæb/	FAITH	*CODA/h	PARSE
☞	[e][hæb]			
	[eh][æb]		*!	
	[e]h[æb]			*!

"

The material discussed in Sections 3 and 4 could be conceived of in this way, and many of the same arguments would go through. As mentioned above, however, I implement phonotactic constraints as part of a 'mixed' rule-and-constraint model, in part because this allows the straightforward depiction of (new) processes as rules. If phonotactic constraints apply to surface forms, as argued in Section 2.1, this means that they can be modelled as applying at the 'end' of phonology, once rules have applied, around the level of surface representation. The precise model that I use has been explicitly advocated in essence in such work as Sommerstein (1974: 72), who assumes that phonotactics apply at a "categoral phonetic" level. The same issues as are in focus in (8) can be implemented in such a model as in (9), so that, to retain Hammond's transcriptions, [ehæb] is possible, but [ebæh] is not, using a series of rule-like syllabification generalisations which apply first, and are then followed by phonotactics (which are set out in a box). I adopt this formalism for the rest of this article.

(9)

	/ehæb/	/ebæh/
assign a Nucleus to sonority peaks in a string	e͡hæ͡b	e͡bæ͡h
gather anything to the left of a Nucleus in an Onset	.e͡.hæ͡b	.e.bæ͡h
gather anything to the right of a Nucleus in a Coda	.e͡.hæ͡b.	.e.bæ͡h.
apply phonotactics: including *CODA-h	[.e͡.hæ͡b.]	*

While there are problems with this kind of model (indeed, we will modify it slightly in the first part of Section 3), the assumption that a key place where phonotactics applies in the phonology is at the end (of a relevant phonological component) will suffice to set out the main issues in this paper: the extent to which such phonotactic constraints are involved in diachrony.

3. The diachrony of phonotactics

Do phonotactics represent units with histories and impacts of their own in diachrony? The phonotactics described for English in Section 2 hold for (some varieties of) present-day English but this has surely not always been the case. Many of the generalisations discussed above can be projected far back into the past: no word of Proto-Germanic begins with /tl/ or /dl/, for example (according to Kroonen 2013), and there have been no major innovations of sibilant deletion, epenthesis or widespread alteration of sibilance in the historical period, which indicates that OCP(SIBILANCE) is also likely not a new constraint. However, the fact that there is variation in terms of the phonotactics of different varieties of English shows that phonotactics can vary and change. For example, all varieties of English used to have [h], but now some do not have it in any phonological environment, which implies that they now have a simple *h constraint, rather than *CODA-h.

There has been some previous work on diachronic phonotactics – as well as discussion of relevant issues in general volumes such as Minkova (2014), some focused work exists, such as

Lutz (1988, 1991), Dziubalska-Kołaczyk (2005) and the work of Ritt and others at Vienna (e.g., Baumann, Ritt and Prömer 2016). Some issues are quite clear to consider, such as the questions of how phonotactics can be innovated into or lost from a language (in OT terms: how do they rise to prominence or fall into insignificance in a language's phonology?). Others are more complex, such as the question of whether already-existing phonotactics can interact with changes which are otherwise independent of the phonotactics. I focus on this latter question in Section 4.

If we look back in the history of English it is clear that phonotactics can 'become active' in a language or 'stop being active'. The sequence [kn] was once possible in English onsets as in *knee*, *knight*, as was [gn] as in *gnat*, *gnaw*, and a way of understanding this is to assume that the constraint forbidding complex onsets with [n] must then have been more complex than the current relevant constraint, given in (6). There is agreement that the initial stops in these sequences were lost during a period between the 14th and 17th century (there is dispute over the precise dating – see Minkova 2014 and Lass 1999). Sequences combining [p, b, t, d] with a nasal have *never* been attested in English, but any constraint involved at earlier stages would have been relatively complex: *ONSET-{p,b,t,d}N or *ONSET-{T–k,g}N is more complex than *ONSET-TN. The innovation of what we could see as (10), which doubtless had intermediate stages and variation, was thus accompanied by what we could see as (11).

(10) k, g > ∅ / __n

(11) *ONSET-{p,b,t,d}N > *ONSET-TN

The change in (11) expresses the diachronic phonotactics that were involved in the change and it would be missing part of the point to ignore it (and to only consider (10)). We might even perceive a pressure to simplify phonological generalisations, as in (11), as part of the explanation for the change.

It would be wrong to think that phonotactic constraints are always assumed (or generalised) when a gap is created by change, however. This is shown in the rounding of low vowels when following [w], which was innovated early in the Early Modern period (Minkova 2014), and which can be understood as something like (12) (following Jespersen 1909, but abstracting away from issues of length).

(12) a > ɔ / w__

This has left many varieties of English with a rounded vowel in words like *wand, quarter, water, warm, wash, wander*. The change was inhibited by a following velar, but was otherwise exceptionless, so we might expect it to have been accompanied by the innovation of a constraint forbidding [wa] (or, more broadly, [w]+low-vowel) sequences. However, although rare, such sequences do not seem to be forbidden in English. For example, when the *WAP phone* was invented (with *WAP* an abbreviation of *Wireless Application Protocol*), *WAP* was pronounced [wap]; the word *quark* is pronounced by British particle physicists as both [kwɔːk] and [kwɑːk], with no hesitation among some to use the latter form; and there seems to have been no problem for speakers to revert the vowel in the form *swam* to an unrounded low vowel "due to the analogy of other preterites: *began, drank, etc.*" (Jespersen 1909: 317). This demonstrates the relevance of prosodic constituents to the formation of phonotactics: the Onset and Nucleus do not form a constituent in the syllable, as shown in (1), so speakers could not construct a phonotactic to forbid [wa] sequences in connection with the innovation of (12). [wa] remained a A-gap, which could be filled, as in the ways just discussed.

Phonotactics can clearly be lost in change. The constraint OCP-ONSET(CORONAL), discussed in (5), seems very robust. It was, however, lost around the Early and/or Late Modern English

period in some varieties, in connection with an innovation which was something like (13) (reinterpreting the formulation in Blevins and Grawunder 2009).

(13) k, g > t, d / # __ l

Forms like [tluːt] *clout* and [dlʊv] *glove* are attested widely in the north of England in the materials gathered for the *Survey of English Dialects* (Orton *et al*. 1962–71), and Blevins and Grawunder (2009) gather a range of evidence for this change, showing that it was widespread in the lexicon in many varieties. This makes sense if we assume that the change involves misperception, as in the model of listener-based changes proposed by Ohala (e.g., 1993), due to the fact that sequences like [tl] and [kl] are acoustically similar – if a learner misperceives a [kl] sequence as [tl] in an utterance of a word like *clout*, they could fix their UR for *clout* with initial /tl/, and would thus not perceive an initial [tl] gap in the lexicon, and so would not assume that the language they are learning has OCP-ONSET(CORONAL), unlike previous generations.

4. Can phonotactic constraints inhibit segmental change?

If we take diachronic phonotactics seriously, the kind of issues discussed in Section 3 clearly arise. My main point in this article, however, is that we should also consider whether phonotactics can have other impacts in historical phonology, as in the article's title. In this section, I consider two case studies from the history of English – the second better known than the first – in which we can identify segmental changes which are basically definable in terms of simple phonological patterning but which have been inhibited in one specific phonotactically-defined phonological environment. Both changes can be seen broadly as cases of 'lenition' and 'reduction' (as long as reduction includes deletion), which are types of changes that do not involve taking material from a phonological context. Rather, they are tied to phonological environments only to the extent that they occur in an implicational hierarchy of environments – they are most likely to be inhibited in some environments, and less likely to be inhibited in others (see Honeybone 2012 for a definition of lenition as 'weakly conditioned' along these lines). This means that the changes are expected in 'weak' prosodic environments and are potentially inhibited in 'strong' environments.

4.1 Mid-Scots θ-debuccalisation

One type of lenition which is well integrated into standard 'lenition trajectories', such as that in (14) (which is adapted from one of the top lines of the trajectory in Lass (1984: 178), in which any movement from left to right counts as a kind of lenition) is the change from Lass's 'stage 3' to 'stage 2'.

(14)

Voiceless stop	→	Affricate	→	Oral fricative	→	Glottal fricative	→	∅
5		4		3		2		1

The input to this change is a fortis oral fricative and the output is [h]. This kind of change is common: Kümmel (2007) lists 11 cases of f > h (including all labial fricatives), 11 of θ > h, 39 of s > h (including other similar sibilant fricatives), 8 of ʃ > h, 28 of x > h, and several other cases which are tied to specific segmental environments. It is often called 'debuccalisation'. As a case of lenition, we can expect debuccalisations of this type to follow the kind of environmental patterning set out in (15), which summarises the basic typology of consonantal lenition environments in which change is expected (see, for example, Ségéral and Scheer 2008 and Balogné Bérces and Honeybone 2012). There is an implicational relationship between these

positions: we would expect a lenition to occur in a 'strong' position only if it also occurs in all 'weak' positions, but we would not be surprised if a lenition occurs in either or both weak positions but not in strong positions.

(15)

STRONG	WEAK	
'initial, onset'	'medial, intervocalic'	'final, coda'
[#__] [C.__]	[V__V]	[__.C] [__#]

This type of environmental patterning is found again and again in lenition (see Honeybone 2008, among much else). It is shown, for example, in the variation that exists across dialects in the patterning of 'Spanish Aspiration', which is a case of s > h. In Southern Peninsula Spanish (e.g., Córdoba Spanish, see Penny 2000), the debuccalisation occurs both syllable- and word-finally (as in *pastel* pa[h]tel 'cake' and *vamos* vamo[h] 'let us go'), but not initially (as in *cemento* [s]emento 'cement'), that is, it occurs only in weak positions. In New Mexican Spanish (see Brown 2005), however, the debuccalisation occurs in all the environments where it is possible in Córdoba, but it also occurs medially (as in *casi* ca[h]i 'amost') and, notably, it also occurs in initial position (as in *cemento* [h]emento 'cement'), that is, it occurs in both strong and weak positions. In no variety does it occur only in strong positions.

Let us now turn to the change in focus here. Like English, Scots basically retained Germanic [θ]. However, in Mid-Scots, what was [θ] in Older Scots now has some variable pronunciations as [h]. This seems to be straightforwardly categorisable as a case of debuccalisation along the lines of (16).

(16) θ > h

The change is recorded in traditional dialect descriptions (e.g., Wilson 1915, Wettstein 1942, Zai 1942) and remnants are found in variationist descriptions of urban varieties of Mid-Scots (such as Johnston 1997, Stuart-Smith and Timmins 2006, Clark and Trousdale 2009). While these remnants are robust, the change did not penetrate throughout Scots and it has now retreated from its greatest extent. There is evidence that it is a non-recent change: it is also found in Ulster Scots (Warren Maguire, p.c.), which indicates that it occurred before Scots was taken to Ulster (the majority of settlement there was in the 17th century). Representative data from the kinds of sources just mentioned is given in (17), which sets out cases of original [θ] in the three basic types of environment given in (15).

(17)

θ-	[h]*ink*	(Glasgow)	'think'
θw-	[hwɛŋz]	(Berwickshire)	'thongs' < OE *þwang*
θr-	[hriː]	(Perthshire)	'three'
-θ-	*no*[h]*ing*	(Glasgow)	'nothing'
-θ-	*any*[h]*ing*	(Glasgow)	'anything'
-θ	*ba*[θ]	(Glasgow)	'bath'
-nθ	*mon*[θ]	(Glasgow)	'month'

The data in (17) imply that this debuccalisation had a peculiar patterning: it occurred in 'strong' initial position ([h]*ink* 'think'), and in the 'weak' intervocalic position (*no*[h]*ing* 'nothing'), but it does not seem to have occurred finally (as in *ba*[θ] 'bath'). The data in (17) is

representative – there is no single attestation in any of the sources I am aware of that gives evidence of θ > h having occurred finally. After a thorough consideration of available data, (Johnston 1997: 507) concludes that "[f]inal /θ/ is retained everywhere". This lenition thus looks very strange and seems to be a clear counterexample to the standard implicational hierarchy of lenition environments set out in (15).

It makes sense, however, if we assume that the change is *not*, in fact, strange but rather behaves like other lenitions, which can either occur only in weak positions (being inhibited in strong positions), *or* across the board (context-free) in both strong and weak positions, but that this patterning has been made opaque due to interaction with the constraint *CODA-h (introduced in (3), (8) and (9)). This is a plausible scenario if we make a few basic assumptions: that *CODA-h has long been part of the phonology of Scots, as in English, given that [h] has never occurred finally in either; that the innovation of (16) involved the introduction of context free θ → h; and that, like all changes, this was initially variable. This gives us the situation set out in (18), which shows the phonology of Mid-Scots after the introduction of θ → h. Given that the process is variable, two derivations are shown for each of two representative words: the left column for both words shows what happens when the process applies and the right column shows what happens when it does not. Importantly: if the process applies to θ in final position (as in *bath*), the result is rendered ungrammatical because it violates *CODA-h and so cannot surface. The alternative derivation, without the application of the debuccalisation θ → h can surface without problem, however. Both derivations for *think* can surface.

(18)

	think		*bath*	
	/θɪnk/	/θɪnk/	/baθ/	/baθ/
θ → h	hɪnk	—	bah	—
*CODA-h	—	—	*	—
	[hɪŋk]	[θɪŋk]	*	[baθ]

There is reason to think that this analysis is on the right lines. 'Spanish Aspiration' does not have the patterning identified for Mid-Scots θ-debuccalisation identified here. Spanish Aspiration *does* occur in codas. This difference between the two cases of the same type of change receives a natural explanation on the assumptions made here. Scots has underlying /h/ which has the kind of restricted surface distribution discussed in (3), (8) and (9), so *CODA-h must be active in its phonology. Spanish does not have /h/, so *CODA-h will not be active in its phonology, and it could therefore not have inhibited the patterning of Spanish Aspiration when it was innovated.

The current Mid-Scots situation, which allows [h] in only a few morphemes (and then only variably), is due to a reanalysis which occurred at a later stage of the history of the varieties. This involved the lexicalisation of [h] into the underlying representations of a few morphemes (which thus have two URs – one with /θ/, as previously, and a novel one with /h/) and the concomitant loss of the θ → h process. This is shown in (19), where *CODA-h has no effect, but is still shown because it still exists in the phonology of the language (to limit the surface occurrence of /h/).

(19)

	think		*bath*
	/θɪnk/	/hɪnk/	/baθ/
*CODA-h	—	—	—
	[θɪŋk]	[hɪŋk]	[baθ]

The situation relevant to our purposes is that in (18). It ties in with the idea (set out in Honeybone 2002, 2003, 2005, 2012) that lenition is not 'caused' by its environment, but can be inhibited by it. Lenition can be prosodically inhibited by strong positions, which is the kind of effect that we see when nothing else intervenes, and which gives rise to the implicational hierarchy of environments in (15). It can also be inhibited by interaction with its melodic environment – for example, medial and final stops rarely lenite in nasal-stop clusters, such as [mb, nt]. What we see in the Scots case considered here is that the effects of lenition can also be inhibited by phonotactics: as modelled here, this is not because the lenition itself is prevented from occurring, but because *CODA-h prevents the form with [h] from escaping from the grammar. If this is right, then in this sense, the phonotactic constraint *CODA-h can be seen to have inhibited the θ > h change. This had a far-reaching effect on the distribution of [h] – even when the surface occurrences of [h] in (18) were reanalysed as being derived from underlying /h/, as in (19), none were assumed to be underlying word-finally, because none had been there at the surface in (18).

4.2 Spätmittelenglischer Schwund of unstressed vowels

My second case study involves vowel reduction in its broadest sense: it is a case of vowel deletion (of syncope). It is not surprising in diachronic phonology if unstressed vowels are reduced and/or deleted. This can involve apocope and syncope, and some example of both are schematised in (20), where the underlined vowels are those which are lost.

(20)	apocope	(post-tonic) syncope
	CV́C<u>V</u> > CV́C	CV́C<u>V</u>C > CV́CC
		CV́C<u>V</u>C<u>V</u>C > CV́CCVC, CV́CVCC

English has lost unstressed vowels in many ways connected with these (and other related) kinds of phenomena. What I discuss here is part of what is described by Luick (1914-40: 534) as *Spätmittelenglischer Schwund* ('late Middle English loss') of such unstressed vowels. I focus on a case of (post-tonic) syncope, of the type schematised in (21), which targeted the last unstressed vowel in a word. As shown in the syllabic representations, this involved the loss of a syllable.

(21)	CV́C<u>V</u>C	>	CV́CC		ό<u>σ</u>	>	ό
	CV́CVC<u>V</u>C	>	CV́CVCC		ό<u>σσ</u>	>	ό<u>σ</u>

Syncope of the type encountered here is not unusual in English, although the precise details of the change in question here are somewhat complex. Luick (1914-40) groups it together with, for example, cases like those schematised and exemplified in (22).

(22)	CV́l<u>V</u>s	>	CV́ls		*elles*	>	*else*
	CV́n<u>V</u>s	>	CV́ns		*hennes*	>	*hence*
					sinnes	>	*since*
					ones	>	*once*

The precise case of *spätmittelenglische syncope* in question here can be described as in (23), which sets out the specific segmental conditioning involved. Luick (1914-1049) assumes that this change is completed by 1500 and Jordan (1974: 141) writes "[f]rom about 1300 – in the North earlier – covered e was ... lost in third syllables", so the beginnings of this change were clearly quite early.

(23) CV́CVz > CV́Cz
CV́CVCVz > CV́CVCz

It has been argued that this change was not widespread (see Minkova 2014). No such syncope occurred in *lettuce, foetus, mattress*, which had CV́CVs, for example, but if we formulate the change as in (23), with a final [z], it may in fact have been exceptionless. It occurred in morphologically simple forms like *Thames* < *Temys, alms* < *almis, adze* < *addis* (which have also undergone 'Middle English final obstruent lenisisation', to produce [z] – see Jespersen 1933 – what Lass (1999) calls 'weak σ voicing'). The change clearly *is* general in inflections, as in the plural '*-es*' (-Vz). Lass (1999: 142) sets out the standard assumptions about the changes involved in the development of the English plural inflection as in (24).

(24) "An idealised history of the early stages of the plurals of *kiss, cat, dog* (the weak vowel represented as /V/) would be:

	kiss	*cat*	*dog*
Early ME input	kis-Vs	kat-Vs	dog-Vs
Weak σ voicing	kis-Vz	kat-Vz	dog-Vz
Weak V deletion	—	*kat-z	dog-z
Voicing assimilation	—	kat-s	— ”

The change in question here is called 'weak V deletion' in (24). The relevant cases are those repeated in (25), augmented to show that not just disyllabic forms were affected. The left-hand, pre-change column indicates forms that existed following 'weak σ voicing' (following Lass's use of [V] to represent the 'weak vowel' and of semi-orthographic forms to fudge the issues surrounding the nature of the tonic vowels), and the right-hand column gives post-syncope forms.

(25) kátVz > kátz [→ káts]
dógVz > dógz
kísVz ... kísVz
máidenVz > máidenz
héavenVz > héavenz
físherVz > físherz
ábbessVz ... ábbessVz

The absence of syncope (indicated by the absence of '>') in *kisses* and *abbesses* is the important thing to note here. It makes sense if we assume (i) that the key change was the introduction of this case of syncope, which fundamentally affected every occurrence of -Vz; (ii) that, like all changes, this was initially variable; and (iii) that it was inhibited due to interaction with OCP(SIBILANCE), which we can reasonably assume has long been part of the phonology of English, given that the fundamental distribution of sibilants has not changed over the recorded history of the language. This is modelled in (26), which shows the variability of the syncopating pre-z *Spätmittelenglischer Schwund* (*SpSchw*) by giving derivations for each of three representative words. This shows synchronic derivations and assumes that the *SpSchw* involved in innovation of a (variable) rule of syncope. As in (18), the left column for each word shows what happens when the process applies and the right column shows what happens when it does not. If the process applies, the result is rendered ungrammatical for *kiss* (and other words which end in a sibilant) because it violates OCP(SIBILANCE), so cannot surface. The alternative derivation, without the application of *SpSchw* can surface without problem in all cases, however. As in (18), both derivations can surface for the other words.

(26)

	kisses		*dogs*		*heavens*	
	/kis+Vz/	/kis+Vz/	/dog+Vz/	/dog+Vz/	/hevən+Vz/	/hevən+Vz/
SpSchw	kisz	—	dogz	—	hevənz	—
OCP(SIB)	*	—	—	—	—	—
	*	[kisVz]	[dogz]	[dogVz]	[hevənz]	[hevənVz]

The current situation in English has been derived by a reanalysis of the inflection's UR along these lines: -Vz > -z, which was accompanied by a 'rule inversion'-type change[15] which involved the loss of the *Spätmittelenglischer Schwund* process and the innovation of a process of epenthesis to break up sibilant-sibilant sequences, of the type $\emptyset \rightarrow V$/ [sib] __ [sib]#, as is often assumed in phonological analyses of present-day English regular pluralisation, such as Hockett (1958: 282) and Jensen (1993: 181). The current situation is shown in (27), which includes OCP(SIBILANCE), because, as in (19), the constraint still exists in the phonology of English even though it has no effect here.

(27)

	kisses	*dogs*	*heavens*
	/kis+z/	/dog+z/	/hevən+z/
epenthesis	kisVz	—	—
OCP(SIB)	—	—	—
	[kisVz]	[dogz]	[hevənz]

What we see in the relevant case of *SpSchw* syncope in (26), is that the phonotactic constraint OCP(SIBILANCE) prevents the syncopated form from escaping from the grammar, in the same way as in (18). If this is right, then the phonotactic constraint can again be seen to have inhibited a phonological change. This again had a far-reaching effect in the phonology of English – when the situation in (26) was reanalysed to give (27), the surface distribution of sibilants was retained, showing the influence of OCP(SIBILANCE), but only by inverting a process which had the effect of innovating a case of epenthesis. The basic conclusion of this section is that the diachronic data from the two case studies only makes sense if we assume that language-specific phonotactic constraints on phonological forms can inhibit otherwise regular changes.

5. Conclusion

Both of the cases of change discussed in Section 4 involve segmental changes which are otherwise general in their phonological environment (or even occur across-the-board) but which have both been inhibited in one specific phonological environment. From their patterning, it does thus seem clear that phonotactics *can* inhibit phonological change – to the extent that the patterning of a change, as visible at the surface level of phonology can be stopped from violating already existing phonotactics. There is no reason to assume that this is *always* the case, however – indeed the discussion in Section 3 concerning the loss of OCP-ONSET(CORONAL) implies that it is not. The points made in this article raise the prospect of an

[15] A reviewer has questioned this description: it can be seen as a case of rule inversion as long as we follow the common analysis of contemporary English plural formation which sees the underlying form of the plural morpheme as /-z/ and the vowel in the plural of sibilant-final words as due to epenthesis (the precise nature of the epenthesised vowel varies from variety to variety, taking in both [ɪ] and [ə]). The rule-inversion analysis is that the alternation between an unstressed vowel and zero exists both pre-change and post-change, and was derived pre-change by the deletion of an underlying vowel, and post-change by the epenthesis of a non-underlying vowel.

intriguing research programme to consider whether we can generalise about when segmental changes are inhibited by phonotactics and when they are not. Clearly the latter must involve more a fundamental reanalysis of a phonological system by those doing the innovation.

I conclude by returning to the issues that began this article. If phonotactic constraints are real phonologically-existing entities which enforce S-gaps in languages as part of a phonological grammar, then we would expect them to be involved in diachrony like any other aspect of phonology. If they are as psychologically real as other kinds of phonological object (such as segments, stress and feet), then we should expect that their effects will be detectable in phonological change. If the material discussed in this article is on the right lines, they are. The diachronic data discussed here provides further evidence that S-gaps are real, and that we need to model them in phonological theory. The ideas developed in Sections 2 and 3 predict that diachronic data of the type discussed in section 4 should exist, and it is only if we make the assumptions set out in Sections 2 and 3 that the data in the case studies make sense. Once again, we can see that the detailed investigation of diachrony can provide evidence for our general understanding of how phonology works.

References

Balogné Bérces, Katalin & Patrick Honeybone. 2012. Splitting 'intervocalic': expanding the typology of lenition environments. *Acta Linguistica Hungarica* 59, 27-48.

Bauer, Laurie. 2015. English phonotactics. *English Language and Linguistics* 19, 437-475.

Baumann, Andreas, Nikolaus Ritt & Christina Prömer. 2016. Diachronic dynamics of Middle English phonotactics provide evidence for analogy effects among lexical and morphonotactic consonant clusters. *Papers in Historical Phonology* 1, 50-75.

Blevins, Juliette & Grawunder, Sven. 2009. *KL > TL sound change in Germanic and elsewhere: descriptions, explanations, and implications. *Linguistic Typology* 13, 267–303.

Boersma, Paul & Silke Hamann. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25, 217-270.

Brown, Esther. 2005. New Mexican Spanish: insight into the variable reduction of 'la ehe inihial' (/s-/). *Hispania* 88, 813-824.

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

Clark, Lynn & Graeme Trousdale. 2009 The role of frequency in phonological change: evidence from TH-fronting in east-central Scotland. *English Language and Linguistics* 13, 33-55.

Coltheart, Max. 1981. The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology* 33A, 497-505.

CUBE = *Current British English searchable transcriptions* = Lindsey, Geoff & Péter Szigetvári. 2013. http://cube.elte.hu/. Consulted May 2018.

Davis, Stuart & Mi-Hui Cho. 2003. The distribution of aspirated stops and /h/ in American English and Korean: an alignment approach with typological implications. *Linguistics* 41, 607-652.

Davis, Stuart & Michael Hammond. 1995. On the status of onglides in American English. Phonology, 12, 159-182.

Dressler, Wolfgang & Katarzyna Dziubalska-Kołaczyk. 2006. Proposing morphonotactics. *Wiener Linguistische Gazette* 73, 1-19.

Dziubalska-Kołaczyk, Katarzyna. 2005. Phonotactics of consonant clusters in the history of English. In: Bertacca, Antonio (ed.) *Historical linguistic studies of spoken English*, 15-34. Pisa: PLUS – Pisana University Press.

Everett-Heath, John. 2014. *The Concise Dictionary of World Place Names*. Third edition; current online version. Oxford: Oxford University Press. DOI: 10.1093/acref/9780191751394.001.0001. Consulted May 2018.
Faust, Noam, Adèle Jatteau & Tobias Scheer. 2018. Two Phonologies. Paper presented at the 26th Manchester Phonology Meeting. Manchester.
Fudge, Erik. 1969. Syllables. *Journal of Linguistics* 5, 253-286.
Gorman, Lyle. 2013. *Generative Phonotactics*. PhD, University of Pennsylvania.
Hammond, Michael. 1999. *The Phonology of English: a prosodic Optimality-Theoretic approach*. Oxford: Oxford University Press.
Harris, John. 1994. *English Sound Structure*. Oxford: Blackwell
Harris, John. 2000. The foot as a segmental domain. Paper presented at the 8th Manchester Phonology Meeting. Manchester.
Hockett, Charles. 1958. *A Course in Modern Linguistics*. New York: Macmillan.
Holder, William. 1669. *The Elements of Speech*. Printed by T. N. for J. Martyn, London.
Honeybone, Patrick. 2002. *Germanic Obstruent Lenition: some mutual implications of theoretical and historical phonology*. PhD thesis, University of Newcastle upon Tyne.
Honeybone, Patrick. 2003. Process inhibition in historical phonology. In Blake, Barry & Kate Burridge (eds.) *Historical Linguistics 2001*, 181-203. Amsterdam: John Benjamins.
Honeybone, Patrick. 2005. Sharing makes us stronger: process inhibition and segmental structure. In Carr, Philip, Jacques Durand & Colin Ewen (eds.) *Headhood, Elements, Specification and Contrastivity: Phonological Papers in Honour of John Anderson*, 167-192. Amsterdam: John Benjamins.
Honeybone, Patrick. 2008. Lenition, weakening and consonantal strength: tracing concepts through the history of phonology. In Brandão de Carvalho, Joaquim, Tobias Scheer & Philippe Ségéral. (eds) *Lenition and Fortition*, 9-93. Berlin: Mouton de Gruyter.
Honeybone, Patrick. 2012. Lenition in English. In Nevalainen, Terttu & Elizabeth Traugott (eds.) *The Oxford Handbook of the History of English*, 773-787. New York: Oxford University Press.
Jensen, John. 1993. *English phonology*. Amsterdam: John Benjamins.
Jespersen, Otto. 1909. *A Modern English Grammar on Historical Principles: Volume 1, Sounds and Spellings*. Heidelberg: C. Winter.
Jespersen, Otto. 1933. Voiced and voiceless fricatives in English. In Otto Jespersen. *Linguistica: selected papers in English French and German*. Copenhagen: Levin and Munksgaard.
Johnston, Paul 1997. Regional Variation. In Jones, Charles (ed.) *The Edinburgh History of the Scots Language*. Edinburgh: Edinburgh University Press.
Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic*. Leiden: Brill. Online here: http://dictionaries.brillonline.com/proto-germanic. Consulted May 2018.
Kruisinga, Etsko. 1943. *The phonetic structure of English words*. Bern: A. Francke.
Kümmel, Martin. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen*. Wiesbaden: Reichert.
Ladefoged, Peter & Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.
Lass, Roger. 1984. *Phonology: an Introduction to Basic Concepts*. Cambridge: Cambridge University Press.
Lass, Roger. 1999. Phonology and morphology. In Lass, Roger (ed.) *Cambridge History of the English Language. Volume 3: 1476–1776*. Cambridge: Cambridge University Press.
Lentz, Tomas & René Kager. 2015. Categorical Phonotactic Knowledge Filters Second Language Input, but Probabilistic Phonotactic Knowledge Can Still Be Acquired. *Language and Speech* 58, 387-413.
Luick, Karl. 1914-1940. Historische Grammatik der englischen Sprache. Stuttgart: Bernhard Tauchnitz. Reprinted (1964) Oxford: Blackwell.

Lutz, Angelika. 1988. On the historical phonotactics of English. In Kastovsky, Dieter & Gero Bauer (eds). *Luick Revisited: Papers Read at the Luick-Symposium at Schloss Liechtenstein, 15.-18.9.1985*, 221-239. Tübingen: Gunter Narr.

Lutz, Angelika. 1991. *Phonotaktisch gesteuerte Konsonantenveränderungen in der Geschichte des Englischen*. Tübingen: Max Niemeyer Verlag.

Minkova, Donka. 2014. *A Historical Phonology of English*. Edinburgh: University of Edinburgh Press.

MRC Psycholinguistic Database = see Coltheart 1981. http://websites.psychology.uwa.edu.au /school/MRCDatabase/uwa_mrc.htm. Consulted May 2018.

OED = *Oxford English Dictionary Online*. Oxford: Oxford University Press. http://www.oed.com/. Consulted May 2018.

Ohala, John. 1993. The phonetics of sound change. In Charles Jones (ed.) *Historical linguistics: problems and perspectives*, 237–278. London: Longman.

Orton, Harold, Michael V. Barry, Wilfried J. Halliday, Philip M. Tilling & Martyn F. Wakelin. 1962–71. *Survey of English dialects*, 4 vols. (each in 3 parts). Leeds: E. J. Arnold.

Paradis, Carole. 2006. The Unnatural /C Ju/ (< Foreign /Cy/) Sequence in Russian Loanwords: A Problem for the Perceptual View. *Lingua* 116, 976-95.

Penny, Ralph. 2000. *Variation and Change in Spanish*. Cambridge: Cambridge University Press.

Peperkamp, Sharon. 2005. A psycholinguistic theory of loanword adaptation. In *Proceedings of the 30th meeting of the Berkeley Linguistics Society*, 341–352. Berkeley: Berkeley Linguistics Society.

Evers, Vincent, Henning Reetz & Aditi Lahiri. 1998. Crosslinguistic Acoustic Categorization of Sibilants Independent of Phonological Status. *Journal of Phonetics* 26, 345-70.

Scheer, Tobias & Philippe Ségéral. 2008. Positional Factors in Lenition and Fortition. In Brandão de Carvalho, Joaquim, Tobias Scheer & Philippe Ségéral. (eds) *Lenition and Fortition*, 131-72. Berlin: Mouton de Gruyter.

Shibatani, Masayoshi. 1973. The role of surface phonetic constraints in generative phonology *Language* 49, 87.

Sommerstein, Alan. 1974. On Phonotactically Motivated Rules. *Journal of Linguistics* 10, 71-94.

Stuart-Smith, Jane & Claire Timmins. 2006. 'Tell her to shut her moof': the role of the lexicon in TH-fronting in Glaswegian. In: Kay et al. (eds.) *The Power of Words*. Amsterdam: Rodopi.

Szigetvári, Péter. 2007. Branching Onsets and Syncope in English. *Language Sciences* 29, 408-25.

Wettstein, Paul. 1942. *The Phonology of a Berwickshire dialect*. Zurich: Bienne.

Wilson, James. 1915. *Lowland Scotch as Spoken in the Lower Strathearn District of Perthshire*. London: Oxford University Press.

Zai, Rudolf. 1942. *The Phonology of the Morebattle Dialect, East Roxburghshire*. Lucerne: Räber.