

**Synchronic stratum-specific rates of application reflect
diachronic change:
Morphosyntactic conditioning of variation in English
/l/-darkening**

DANIELLE TURTON

Newcastle University

Abstract

Phonological processes that exhibit morphosyntactic sensitivity can provide evidence of historical processes which have ascended through the grammar over time. English /l/-darkening shows such effects. Although syllable-based accounts state that light [l] occurs in onsets (e.g. *light*) and dark [ɫ] in codas (e.g. *dull*), several studies report overapplication of darkening to onset /l/ in certain morphosyntactically defined positions: e.g. word-finally in phrases such as *heal it*, and stem-finally before a suffix in words such as *heal-ing*. Although many phonological theories attempt to account for such opacity, they cannot adequately account for the potential variability in application alongside this.

The present paper explores these ideas through modelling data on /l/-darkening in English taken from Hayes's (2000) Optimality Theoretic study. It is argued that a combined Stochastic Stratal OT approach to the data is an improvement over a parallel stochastic model (e.g. Boersma & Hayes 2001) because it avoids fixed innate constraint rankings, which are required to prevent the prediction of impossible grammars. Moreover, it is shown that observations about the diachronic life cycle of phonological processes enable us to deduce quantitative predictions about rates: the process should apply with lower frequency in smaller morphosyntactic domains.

1 Introduction

Since the advent of Optimality Theory (Prince & Smolensky 1993; henceforth OT), numerous methods have been devised to describe phonological variation within a constraint-based framework (Anttila 1997, 2007; Coetzee 2004; Coetzee & Pater 2011). One of the most successful is Stochastic OT (Boersma 1997). In this theory, optimality-theoretic constraints have

ranges over a continuum of strictness, rather than fixed points on a hierarchy: since these ranges may overlap, variation will arise. One notable advantage of Stochastic OT is its association with the Gradual Learning Algorithm (GLA). When exposed to a training corpus, the GLA has been shown to select a grammar which, in a large range of cases, successfully generates the relative frequencies of variants in the corpus.

In their seminal paper testing the potential of the GLA, Boersma & Hayes (2001) demonstrated the advantages of Stochastic OT by modelling results from several empirical studies, notably Hayes's (2000) data on English /l/-darkening. Hayes demonstrated that the process of /l/-darkening in English is morphosyntactically conditioned, yielding higher frequencies of dark [ɫ] in complex *heal-ing* than in monomorphemic *Hay-ley*. Such alternations raise challenges for the study of variation as, in addition to a model of variable processes, they require a theory of the morphosyntax-phonology interface. To this end, Hayes adopts Output-Output Correspondence (OOC), a theory incorporating constraints that demand identity between morphologically related surface forms (Benua 1995, 1997a; Kenstowicz 1996). However, OOC has crucial drawbacks, which become apparent in Hayes's (2000) own approach to /l/-darkening. Bermúdez-Otero (2011:2043) shows that the OOC constraints used by Hayes can generate an impossible dialect in which darkening overapplies at stem-suffix boundaries (e.g. *hea[ɫ]-ing*) but not at word boundaries (e.g. *hea[l] it*). These problems, discussed further in section 2.2, warrant the exploration of alternative approaches to morphosyntactically conditioned phonological variation.

This paper explores the possibility of combining Stochastic OT with a different theory of the morphosyntax-phonology interface: Stratal OT (Bermúdez-Otero 1999, 2011, Forthcoming; Kiparsky 2000). In Stratal OT, phonological processes apply cyclically over a hierarchy of stem-level, word-level, and phrase-level domains, with each domain subject to its own stratum-specific OT grammar. Stratal OT predicts without stipulation that, if darkening overapplies at stem-suffix boundaries, it will also overapply at word boundaries.

Using the data on English /l/ from Hayes (2000) and Boersma & Hayes (2001), this paper will show that a Stochastic Stratal OT (SSOT) analysis can match the patterns of variation noted by Boersma & Hayes with as much accuracy as a parallel model based on OOC. Nevertheless, the crucial advantage of a stratal analysis is that it is capable of correctly generating the patterns of morphosyntactically driven variation in /l/-darkening without simultaneously predicting the existence of impossible dialects by factorial typology.

This application of a stratal-cyclic theory of phonology to morphosyntactically conditioned variation develops ideas found in previous work by Guy (1991a,b). In an influential series of papers, Guy combined the Labovian idea of variable rewrite rules (Labov 1969) with Lexical Phonology (Kiparsky 1982) in order to explain morphophonological patterns found in τ/D -deletion. Guy found that there was more deletion of the final consonant in monomorphemic *mist* than in the past-tense form *miss-ed*. Using a cyclic framework, Guy argued that monomorphemes such as *mist* show the highest frequency of rule application because the conditions of τ/D -deletion are met at more levels of the derivation (see section 2.3). However, Guy assumed *a priori* that variable rules have the same probability of application at all derivational levels. This is not what would be predicted given a diachronic approach that takes into consideration the life cycle of phonological processes (Bermúdez-Otero 1999, 2015; Bermúdez-Otero & Trousdale 2012; Turton 2014a; Ramsammy 2015). Since phonological processes become increasingly integrated with morphosyntactic structure as they age, an analysis acknowledging the life cycle would instead predict that the rate of application of a variable phonological process also changes over time. Specifically, when an innovative phonetically gradient process stabilises as a categorical phonological process, it initially applies at the widest domain of the phrase level and can see across word boundaries. Over time, a rule may advance to the word level, and finally the stem level. This means that a rule which currently applies at the stem level progressed by previously applying at the word and phrase levels. It follows that phrase-level application of a variable phonological process will be higher than word-level application because the rule has been active at the phrase level for a longer period of time. This prediction (discussed in detail in section 2.3) does not follow from Guy's analysis, which instead stipulates equal rates of application at each cyclic level.

In view of the problems associated with OOC, and of the predictions arising from the life cycle of phonological processes, the goal of this paper is to provide a reanalysis of Boersma & Hayes's data by calculating *cycle-specific* rates of /l/-darkening. We will see that an SSOT approach successfully matches the frequencies modelled in Boersma & Hayes (2001). Finally, it will be shown that frequencies calculated at each individual stratum corroborate the predictions made by the life cycle of phonological processes.

2 Background

2.1 Morphosyntactic conditions on /l/-darkening

Traditional descriptions of English /l/-darkening posit two discrete allophonic categories: light [l] and dark [ɫ] (Chomsky & Halle 1968; Halle & Mohanan 1985). Although the categorical nature of this distinction has been challenged, notably by Sproat & Fujimura (1993), section 2.5 below discusses the evidence supporting /l/-darkening as a discrete process. Acoustically, dark [ɫ] is characterised by a close proximity between the first and second formants, in comparison to light [l] which has a relatively high F2 and a low F1 (Ladefoged & Maddieson 1996; Carter 2003; Hawkins 2004; Recasens & Espinosa 2005:3; Gick et al. 2006; Carter & Local 2007). The articulatory correlates include tongue dorsum backing or lowering (Lee-Kim et al. 2013; Turton 2014a,b, Forthcoming) and the relative phasing of the coronal and dorsal gestures (Sproat & Fujimura 1993; Huffman 1997; Lin 2011): there is coronal lead in light [l] and coronal lag in dark [ɫ].

In syllable-based analyses of the process, it is noted that light [l] occurs in canonical onsets (*like, love*) and dark [ɫ] in canonical codas (*pool, dull*; Giles & Moll 1975; Giegerich 1992; Roach et al. 2006). However, in some dialects of English the process has been found to overapply when /l/ surfaces in the onset, yielding dark [ɫ] word-finally in phrases such as *heal it*, and stem-finally before a suffix in words such as *heal-ing* (Jensen 1993; Olive et al. 1993; Carter 2003; Bermúdez-Otero 2007; Turton 2014a,b).¹ These instances of opaque overapplication demonstrate that /l/-darkening is sensitive to morphosyntactic structure, a fact which Boersma & Hayes (2001) seek to incorporate into their model.

2.2 Avoiding unattested dialects

To account for the unexpected presence of onset-position dark [ɫ]s in morphosyntactically complex items such as *healing* and *heal it*, Hayes (2000) relies on the fact that, in the citation form *heal*, the dark [ɫ] occurs transparently. Accordingly, Hayes invokes two OOC constraints which require

¹ I assume here that English has full resyllabification of word-final and stem-final prevocalic consonants into the onset of the following word or suffix. Processes such as /l/-vocalisation and /r/-loss in non-rhotic dialects, which target segments in canonical coda positions, support this assumption: they crucially do not apply to word-final prevocalic consonants (Bermúdez-Otero 2011: 2039). Minkova (2003) provides persuasive evidence to date the rise of resyllabification to the Middle English period.

faithfulness between the derived output form and the output base. IDENT-OO(Phrasal) monitors the correspondence between the base and the surface form of a word in complex phrasal environments, whilst IDENT-OO (Morphological) penalises differences between the base and the surface form of a stem in morphologically complex words (see fig. 1).

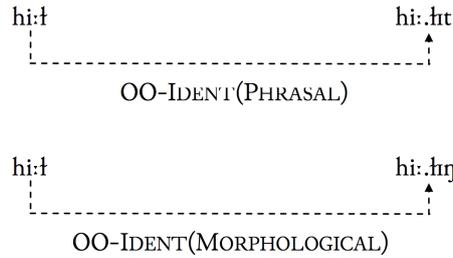


Figure 1: From Bermúdez-Otero (2011:2043)

By factorial typology, however, Hayes's constraints predict unattested darkening patterns in morphologically complex words. If IDENT-OO(Morphological) is ranked higher than IDENT-OO(Phrasal), then a problematic grammar arises, as table 1 demonstrates. This ranking produces a dialect with opaque dark [ɫ] in the onset in *healing* [hi:tŋ], by correspondence with the base *heal* [hi:t], but with transparent light [l] in *heal it* [hi:lt]: i.e. [hi:tŋ, hi:lt, hi:t]. This is henceforth referred to as the *[hi:tŋ, hi:lt] dialect. It is generally the case that if a phonological process applies opaquely at stem-suffix boundaries, it will also apply opaquely at word boundaries. That is, the existence of a dark [ɫ] in *healing* requires the /l/ in *heal it* to be dark.

Hayes (2000: 102) acknowledges that grammars such as the *[hi:tŋ, hi:lt] dialect are unattested and in order to avoid them suggests that Paradigm Uniformity constraints should be *a priori* stricter for higher in phrase than in words, ranked in UG. By stipulating the innate ranking in (1), Hayes's analysis avoids the unattested *[hi:tŋ, hi:lt] dialect:

- (1) OOC-PHRASAL \gg OOC-MORPHOLOGICAL

However, if the two OOC constraints are truly independent they should be freely rankable. Thus, Hayes's solution has been criticised by Bermúdez-Otero (2011), who points out that this innate ranking stipulation is unnecessary in a cyclic framework. Bermúdez-Otero argues that the *[hi:tŋ, hi:lt] dialect is not only unattested, it is impossible by the Russian Doll Theorem:

| | | IDENT-OO (Morphological) | PREVOCALIC /l/ IS LIGHT | IDENT-OO (Phrasal) | /l/ IS DARK |
|----------------|----|-----------------------------|----------------------------|-----------------------|-------------|
| <i>healing</i> | a. | hi:lɪŋ | | | * |
| | b. | hi:tɪŋ | * | | |
| <i>heal it</i> | a. | hi:lɪt | | * | * |
| | b. | hi:tɪt | *! | | |
| <i>heal</i> | a. | hi:l | | | *! |
| | b. | hi:t | | | |

Table 1: Impossible dialect predicted by Hayes's constraints in factorial typology

(2) The Russian Doll Theorem

Let there be the nested cyclic domains $[\gamma \dots [\beta \dots [\alpha \dots] \dots] \dots]$. If a phonological process p is opaque in β because its domain is α , then p is opaque in γ .

(Bermúdez-Otero 2011: 2026)

Bermúdez-Otero describes the Russian Doll Theorem as an entailment of cyclic theory which has not been formally articulated before, most likely due to its obviousness. However, Hayes's study shows that, without the innate ranking stipulation in (1), OOC can easily generate outputs that violate the Russian Doll Theorem. Bermúdez-Otero explains that, in a stratal framework, this stipulative ranking is unnecessary because the work of OOC is done by faithfulness between cycles. An /l/ which is darkened in a given stratum remains dark in subsequent cycles.² Crucially, what remains to be demonstrated is how a cyclic framework can accurately model the variability of the /l/-darkening processes in the same way as parallel Stochastic OT. As we shall see in section 2.3 below, the key advantage of SSOT is that it correctly generates the patterns of variation in /l/-darkening without stipulating any innate ranking of OOC-constraints.

2.3 Modelling variation in Stratal OT

It has been argued that Stratal OT performs better than OOC as a theory of the morphosyntax-phonology interface for /l/-darkening, as it does not generate the problematic *[hr:ħŋ, hr:l it] dialect. The challenge now is to test whether an approach incorporating both Stochastic and Stratal OT proves as successful as Boersma & Hayes (2001) in matching the relative frequencies of light and dark /l/.

A framework in which /l/-darkening applies variably over several cycles crucially predicts that the probability of a dark realisation on the surface will depend on how many cycles in the derivation meet the conditions for darkening (i.e. how many cycles in the derivation place /l/ in coda position). In other words, the model predicts that, for a given phonological environment, the greater the number of cycles in which the conditions for darkening are met, the higher the number of darkened tokens. This idea is schematised in table 2, which shows that, in the word

² Ramsammy (2011) diagnoses the same problem in Baković's (2001) OOC account of Spanish nasal velarisation. In Baković's (2001) analysis, compliance with the Russian Doll Theorem is stipulated; Ramsammy provides a Stratal OT account in which it is predicted.

heal, the prepausal /l/ is in the coda in the stem-level, word-level, and phrase-level cycles. Compare this to *healing*: the /l/ is in the coda only at the stem level, as it becomes resyllabified into onset position at the word level when the suffix *-ing* becomes visible, remaining in the onset at the phrase level. Therefore, the number of cycles in which the conditions for darkening are met is greater in *heal* than in *heal it*, and greater in *heal it* than in *healing*. This means that dark [ɫ] will occur more frequently in tokens of *heal* than in *heal it* or *healing* as the conditions for darkening are met in all three cycles.

| Environment | Stem level | Word level | Phrase level | /l/ in coda? |
|-----------------|------------|------------|--------------|--------------|
| <i>heal</i> | [hi:l] | [hi:l] | [hi:l] | ✓✓✓ |
| <i>heal it</i> | [hi:l] | [hi:l] | [hi:l it] | ✓✓ |
| <i>heal-ing</i> | [hi:l] | [hi:lɪŋ] | [hi:lɪŋ] | ✓ |

Table 2: Number of cycles meeting the conditions for darkening

As previously pointed out, Guy (1991a,b) was the first to note that words which meet the conditions of application of a variable rule at more cyclic levels will show a higher rate of application overall. Thus, Guy's model captured the observation that there is more T/D-deletion in monomorphemic base forms such as *mist* (cf. *heal* above) than in morphologically derived forms, such as past tense *missed* (cf. *healing* above), as shown in table 3. However, in his cyclic variation model, Guy assumed

| Environment | Level 1 | Level 2 | T/D in complex coda? |
|----------------|---------|---------|----------------------|
| <i>mist</i> | [mist] | [mist] | ✓✓ |
| <i>miss-ed</i> | [mis] | [mis-t] | ✓ |

Table 3: Derivation of T/D-deletion, based on Guy (1991b,a)

equal application rates of the variable rule at all cycles, contra the predictions from the life cycle of phonological processes (Bermúdez-Otero 1999, 2011; Harris 1989; McMahan 2000). As outlined in the introduction, a diachronic analysis that pays due regard to the life cycle would predict the highest rates of application at the phrase level, followed by the word and stem levels respectively. This follows from the observation that young rules initially apply at the largest morphosyntactic domain (i.e. the phrase level) where the entire utterance is visible. Input restructuring results in the rule climbing up the hierarchy of levels: as a rule ages, its

domain of application shrinks (Bermúdez-Otero 2011; Bermúdez-Otero & Trousdale 2012). Over time, the phrase-level rule may climb up to the word level and eventually the stem level. Let us assume that each step of domain narrowing in the life cycle of a phonological process involves a transitional variation period. During this period, the frequency of application of the process in its new domain increases over historical time, possibly along the familiar S-curve pattern shown in fig. 2.

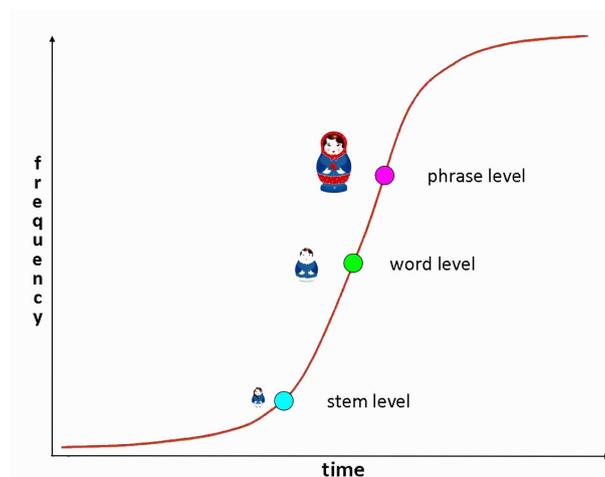


Figure 2: The predictions of the life cycle in a model with variable processes

Considering that a rule enters the phrase level first, it will apply at a higher rate than in the more embedded morphosyntactic domains of the word and stem levels, due to the fact it has been active at that level for a longer period of time. That is, the phrase-level frequency of a process will always be higher than (or equal to) its frequency at the word level, which in turn will be higher than (or equal to) the frequency at the stem level. Regardless of which level a particular phonological process happens to be currently active at, it will have entered the grammar at the phrase level, and so it will have had more time to increase in frequency there (possibly along the S-Curve pattern). Considering the predictions of the life cycle alongside the implications in fig. 2, the following corollary can be added to the Russian Doll theorem:

(3) **The Variation Corollary of the Russian Doll Theorem**³

If a phonological process π shows a rate of application x in a small embedded domain α , then π will apply at a rate equal to or greater than x in a wider cyclic domain β .

The corollary in (3) expresses the idea behind fig. 2: a variable phonological process will apply at its highest rate in the largest morphosyntactic domain, and at its lowest rate in the smallest, most embedded domain. This means that, in contrast with Guy's stipulation of equal rates of application across lexical strata, the present method considers a model in which rates of application are calculated individually from empirical data for each cyclic level. This will enable the predictions from the Variation Corollary of the Russian Doll Theorem to be tested. Thus, the present approach does not start from an assumption that either analysis is correct, but will propose a method to determine which of the two does a better job of fitting the empirical data.

To determine the contribution of each level to overall rates of /l/-darkening on the surface, the procedure used here involves comparisons between different expressions in which darkening is applicable. By selecting a form in which darkening applies, for example, in only one cycle, and comparing it with forms in which darkening occurs in two or three cycles, it is possible to isolate cycle-specific effects, the calculations of which are explained in detail in section 3.1. For example, referring back to table 2, the difference in darkening rates between environments *heal* and *heal it* is one phrase-level cycle of darkening, making it possible to isolate that particular rate of application.

The question that remains is whether the output of an SSOT grammar succeeds in matching the frequencies in Boersma & Hayes (2001) as well as their own stochastic approach with OOC. If so, an SSOT analysis is an improvement on Boersma & Hayes (2001) to the extent that it removes the need for the innate ranking stipulation in (1), and an improvement on Guy (1991a,b) in that it does not assume equal rates of application at each level, but rather seeks to learn the actual individual rates. Thus far, nobody has attempted to work out the contribution of each cyclic level to the surface application rate of a variable phonological process, and this paper develops a method which attempts to do this for the first time.

³ (3) is expected to hold true only in the absence of innovations which enter the grammar from below leading to the loss of π .

2.4 Frequencies of light and dark /l/

The data used in this paper is taken from Hayes's (2000) study on American English /l/-darkening. Hayes (2000) gathered well-formedness ratings of light and dark /l/ in a variety of environments. Boersma & Hayes (2001:82) used a sigmoid transformation to convert these well-formedness ratings into the conjectured frequencies illustrated in table 4.

| Environment type | Conjectured frequency of light [l] % |
|------------------|--------------------------------------|
| <i>light</i> | 99.956 |
| <i>free-ly</i> | 94.53 |
| <i>yellow</i> | 76.69 |
| <i>mail-er</i> | 16.67 |
| <i>mail it</i> | 0.49 |
| <i>bell</i> | 0.0011 |

Table 4: Adapted from Boersma & Hayes (2001:74)

In the original study by Hayes (2000), 10 speakers of American English heard light and dark pronunciations of /l/ in several representative forms, such as those in table 4. They were then asked to rank the acceptability of each pronunciation. Hayes did not directly control the phonetic quality of /l/ in the stimuli, but rather that of the preceding vowel. Hayes (2000:96) justified this procedure on the grounds that that front or high vowels and 'true diphthongs' preceding dark [ɫ] have a schwa off-glide. Therefore, the presence of breaking in the preceding vowel was taken as a proxy for darkening. For example, in a word such as *pool*, dark tokens would be realised as [pu:əɫ] and light tokens as [pu:l]. The acceptability of a broken vowel was taken to equal the acceptability of a dark [ɫ].

Note that Hayes's speakers display a more complex pattern of darkening than was found in some previous studies, such as Sproat & Fujimura (1993). Firstly, they prefer dark /l/ in *mail-er*, compared to Sproat & Fujimura's speakers who have light /l/ in this phonological context. One possible explanation for this is that Hayes's informants represent a variety of American English which is more advanced in its use of dark [ɫ], in that coda-based darkening has climbed up to the stem level. Additionally, Hayes's speakers accept darkening in monomorphemes such as *yellow*, *mellow*, *Hayley*. In these examples, the /l/ is in the onset at all three levels of the cyclic derivation. This suggests the presence of a more aggressive form of darkening that targets foot-medial onset positions.

Table 5 (adapted from Bermúdez-Otero 2007) outlines the /l/-darkening situation in four dialects of English, with Hayes's speakers falling under Stage 4.⁴ As suggested above, this stage exhibits a foot-based process of darkening, whereby the /l/ darkens not only in the coda, but also anywhere outside of foot-initial onset (see also Jensen 1993; Carter 2003; Carter & Local 2007). This is not surprising, given that many lenition processes in English target consonants not only in codas, but also in foot-medial onsets, such as American English /t/-flapping (Kiparsky 1979), British English /t/-glottalling (Harris & Kaye 1990) and Conservative RP /r/-tapping (Jensen 2000).

| | <i>light</i> | <i>yellow</i> | <i>healing</i> | <i>heal it</i> | <i>heal</i> |
|---------|--------------|---------------|----------------|----------------|-------------|
| Stage 1 | [l] | [l] | [l] | [l] | [ɫ] |
| Stage 2 | [l] | [l] | [ɫ] | [ɫ] | [ɫ] |
| Stage 3 | [l] | [ɫ] | [ɫ] | [ɫ] | [ɫ] |
| Stage 4 | [l] | [ɫ] | [ɫ] | [ɫ] | [ɫ] |

Table 5: Dialects at different stages show how /l/-darkening advances through the phonology

This provides evidence for the argument that the life cycle is governed by both a grammatical and a prosodic force. The case for cyclic domains has been discussed in detail above, but Bermúdez-Otero (2010) suggests that we also need to consider the prosodic hierarchy. Following previous discussions of sound change in generative phonology (Vennemann 1972; Kiparsky 1998), Bermúdez-Otero draws attention to the phenomenon of 'rule generalisation', whereby a phonological process first applies in a relatively narrow phonological environment and progresses to more inclusive environments over time. In this respect, Hayes's /l/-darkening data exhibits similarities with the /r/-vocalisation patterns discussed in Harris (2006). The latter provides evidence of a 'broad non-rhotic' dialect of American English where /r/ vocalises not only in *car* and *farm*, but also in *very* and *sheriff*. However, the /r/ is retained in words when a stressed vowel follows within the same word, as in *terrain* and *carouse*, showing that the process of /r/-vocalisation targets /r/s which are non-foot initial. Thus, the effects of the rule generalisation are reflected in

⁴ Stage 1 reflects RP; Stage 2, the American English dialects in Sproat & Fujimura (1993). For Stage 3 see Olive et al. (1993). For Stage 4 see Hayes for American English dialects and Carter (2003) for Northern British English dialects.

changes in the prosodic domain of the process, which first targets weak syllabic positions (codas), and later advances to weak positions in the foot. The present analysis will consider the effects of both coda-based and foot-based /l/-darkening at each level of the cycle.

2.5 Categorical vs. gradient approaches

Although light and dark /l/ have long been considered discrete allophones of the same phoneme (Chomsky & Halle 1968; Halle & Mohanan 1985), there is continued debate in the literature as to whether the distinction is truly categorical. Sproat & Fujimura (1993) concluded from their own X-ray microbeam data that differences in /l/ realisation do not reflect a categorical distinction, but rather a phonetic continuum, crucially correlated with the duration of the pre-boundary rime (the longer the pre-boundary rime, the darker the pre-boundary intervocalic /l/). Hayes (2000:95), who assumes darkening to have a categorical component, rejects Sproat & Fujimura's duration-driven analysis, arguing that their data provide evidence for two phonetic categories partially obscured by free variation. Studies in recent years have seen continued disagreement over the categorical nature of /l/-darkening.

Bermúdez-Otero (2007b:4) offers several arguments against a purely phonetic approach, paying particular attention to the morphosyntactic conditioning of /l/. Evidence for this conditioning can be seen in table 4, and is exemplified by near homophones showing over 60% higher rates of light /l/ in *Norman Mailer* than in complex *mail-er* (Boersma & Hayes 2001). In a modular feedforward architecture of grammar, phonetics is insensitive to morpheme boundaries (Myers 2000:263; Bermúdez-Otero & Luís 2009) and so differences such as these are taken as evidence of two categories.

Furthermore, Bermúdez-Otero & Trousdale argue that Sproat & Fujimura's data *are* compatible with a categorical analysis. As shown in table 6, Sproat & Fujimura find that, for /l/s in initial and intervocalic position, the coronal gesture precedes the dorsal gesture, whereas in word-final position (whether prevocalic or not) the dorsal gesture precedes the coronal. Visual inspection of their plot (1993:303) shows two separate clusters on the coronal delay dimension, suggesting a bimodal distribution, which is commonly argued to represent the existence of separate categories (Bermúdez-Otero & Trousdale 2012: 6). This demonstration is taken a step further by Turton (Forthcoming), who shows that accounting for two categories in Sproat & Fujimura's data provides a better fit when running simple linear models.

| Environment | Example | Gesture | /l/ |
|---------------------------|---------------------|------------------|-----|
| Initial | <i>Likkovsky</i> | Coronal gesture | |
| Intervocalic | <i>Mr Beelik</i> | precedes dorsal | [l] |
| Intervocalic pre-boundary | <i>beel-ing</i> | | |
| Word-final prevocalic | <i>Beel equates</i> | Dorsal gesture | |
| Final | <i>Neal</i> | precedes coronal | [ɫ] |

Table 6: Adapted from the results of Sproat & Fujimura (1993)

In fact, recent statistical, acoustic and articulatory investigation into English /l/-darkening demonstrates that both categorical and gradient processes exist. Using ultrasound tongue imaging data from various dialects of English spoken in the UK, Turton (2014a, Forthcoming) shows that, indeed, some varieties do not seem to have a clear categorical difference (e.g. Manchester, Belfast) whereas others do (e.g. Received Pronunciation, London). Moreover, Turton shows that those with a categorical distribution exhibit evidence of gradient effects of duration, overlaid on darker tokens. This result is argued to provide evidence of rule scattering (Bermúdez-Otero 2015), where two rules which are diachronically related (i.e. the phonetically gradient rule of longer duration in darker /l/s and the categorical process of /l/-darkening) coexist in the synchronic grammar.

This is not the first work to demonstrate that both categorical and gradient effects can both be at work in English /l/-darkening. The phonetic data presented by Yuan & Liberman (2009, 2011) raise significant challenges to the claim that light and dark /l/ are a single phonological entity. Contrary to Sproat & Fujimura, they found that final /l/ was always dark, even in very short rimes, and that clear [l] showed no correlation with duration.

Such data support the idea that darkening originally occurred as a duration-sensitive gradient phonetic process which has, over time, been reanalysed by speakers and phonologised as a duration-insensitive categorical process. This means that the original duration-sensitive gradient process of phonetic implementation coexists in the grammar on top of the newer morphosyntactically conditioned categorical process. Considering both the empirical evidence and the theoretical debate, this paper also assumes that /l/ has a categorical component, suitable for analysis under Stratal OT.

Yuan & Liberman assume a different method of syllabification to the current paper, classifying intervocalic segments as ambisyllabic. Note that this approach accounts for their data only and not the results of other /l/-darkening studies such as Sproat & Fujimura (1993). Ambisyllabicity has been posited to account for phonological patterns found in all intervocalic positions, both within a word *and* across word-boundaries, such as American English /t/-flapping (Kahn 1976). However, ambisyllabicity has been shown to provide an inconsistent account of allophony in English (see Kiparsky 1979; Jensen 2000; Harris 2003; Bermúdez-Otero 2007, 2011). Moreover, Bermúdez-Otero (2007) provides an ambisyllabicity paradox specific to /l/-darkening. Using Sproat & Fujimura's data, he points out a categorical discrepancy between the two supposed ambisyllabic positions. As table 6 above summarises, for a word-final prevocalic /l/ (such as in *Beel equates*) the dorsal gesture precedes the coronal, but for a word-medial intervocalic /l/ (e.g. *Beelik*) the coronal gesture precedes the dorsal gesture. Furthermore, an ambisyllabic analysis is problematic for the Hayes dataset, where we see higher frequencies of darkening in word-final prevocalic position than word-medial intervocalic position. Bermúdez-Otero uses this evidence, alongside examples from other phonological processes, to argue that word-final prevocalic consonants are never ambisyllabic in English.

3 The new model

This section focusses on the new method for calculating individual rates of application of /l/-darkening at each level of the cyclic derivation. Note, however, that the results can only be regarded as proof of principle. Firstly, there are limitations to Hayes's data (and by implication to Boersma & Hayes's conjectured frequencies), which are twofold: they are based on well-formedness judgements, rather than direct observations and measurements of frequencies of light and dark /l/, *and* there are crucial environments missing from Hayes's analysis, such as forms where /l/ is followed by a stressed vowel after a stem-suffix or word boundary (e.g. *mail-ee*, *mail Ann*).

Additionally, the procedure described here does not consist of an unsupervised learning algorithm projecting a full stratal grammar from surface data. Darkening rates are calculated by hand before being fed to the GLA in separate cycles, and therefore I do not claim to provide a learning algorithm for SSOT grammars. The aim of this paper, however, is different: namely, to demonstrate the existence of an SSOT grammar which can model the data as well as Boersma & Hayes, without the innate stipula-

tion of the constraint ranking in (1), and allowing for cycle-specific rates of darkening. With the data that is available, it is perfectly feasible to calculate rates of application at each cyclic level in order to demonstrate the proposed model and show the existence of an SSOT grammar.

In line with the discussion in section 2.3, the calculations that follow will turn out to produce results that comply with the predictions made by the life cycle of phonological processes and the Variation Corollary of the Russian Doll Theorem. That is, a phonological process operating at all three levels of the grammar will show higher rates of application at the phrase level than the word level, and at the word level than the stem level. The life cycle predicts that processes become more embedded as they age; therefore, the stem-level process, which has not been active for as long as the ones at the word and phrase levels, applies at a lower rate (see fig. 2). To carry out the calculations, environments containing /l/ in different positions will be compared in an effort to isolate cycle-specific frequencies.

3.1 Separating coda and foot-based processes

The first issue is to disentangle coda-based and foot-based darkening. As discussed in section 2.4, the /l/ in words such as *yellow* and *Hayley* is never in the coda at any level, demonstrating that /l/ is also susceptible to darkening when non-initial in the foot. This is a concern when working out frequencies in environments such as *mail it*, where the /l/ shows darkening from both coda-based and foot-based processes. These processes need to be separated in order to calculate cycle-specific values of darkening.

Consider a monomorphemic word such as *yellow*. In Stratal OT, this word is subject to precisely three cycles in the phonology; at the stem level, the word level and the phrase level. This /l/ is not in coda position in any of the three cycles, but is in foot-medial position (that is, outside the foot-initial onset) in all three cycles. Therefore, the figure given by Boersma & Hayes for this environment must be the frequency of light [l] after three processes of foot-based darkening only, with no effect from coda-based darkening.

(4)

$$R_{\text{yellow}} = F_s \times F_w \times F_p = 0.766917$$

The equation in (4) states that ‘the retention of light /l/ in *yellow* is equal to the product of foot-based retention at the stem, word and phrase

levels.’ Note that calculations are based on the retention of light [l], not the frequency of derivation of dark [ɫ]. This is in line with Boersma and Hayes’s presentation, and with Guy’s (1991a; 1991b) methods of calculating the percentage of unlenited forms as a product of three cycles. Thus, F_x (where x refers to either the stem, word or phrase level) is a factor between 0 and 1 expressing the likelihood that a light [l] in the input to level x will remain light if it meets the conditions for foot-based darkening. In order for a token to remain light, it must pass through the three levels without darkening, thus the factors are multiplied together.

As discussed in section 2.3, Guy’s analysis of T/D-deletion stipulated that the same retention factor R would apply at all three levels in the grammar. This meant that the overall surface retention rate was the result of applying R three times: i.e. $R \times R \times R = R^3$. This is why his model is described as ‘exponential’. Given these assumptions, Guy calculated R simply by taking the cube root of the surface retention rate. By the same token, a Guyan approach to foot-based /l/-darkening would take the cube root of the figure in 4. The aim of the present paper is to improve on this, and to derive cycle-specific retention rates from empirical data. In this case, however, we come up against the limitations of Hayes’s data: the lack of crucial forms like *mail-ee* and *mail Ann* impedes the calculation of level-specific rates of foot-based darkening. This problem will be revisited below after consideration of the coda-based process.

3.2 Coda-based darkening

Consider *mail-er*, which undergoes foot-based darkening at all three levels of the derivation but undergoes coda-based darkening at the stem level only; the suffix *-er* does not become visible until the word level, where the /l/ is resyllabified into the onset (see table 7).

| Domain | $[_{PL}[_{WL}[_{SL} \text{meil}] \text{ər}]]$ | /l/ non-foot-initial? | /l/ in coda? |
|--------------|--|-----------------------|--------------|
| stem level | $[_{\Sigma}[_{\sigma} \text{meil}]]$ | yes | yes |
| word level | $[_{\Sigma}[_{\sigma} \text{meil}][_{\sigma} \text{ləɹ}]]$ | yes | no |
| phrase level | $[_{\Sigma}[_{\sigma} \text{meil}][_{\sigma} \text{ləɹ}]]$ | yes | no |

Table 7: Stratal-cyclic derivation of *mail-er*

Using the same notation as in (4), with C_x referring to the retention rate for coda-based darkening at level x , we can now express the retention rate of light [l] in *mail-er*. Dividing the frequencies of light [l] in *mail-er* by

that in *yellow* isolates the rate of coda-based retention at the stem level, as shown in (5):

(5)

$$\begin{aligned}\frac{R_{\text{mail-er}}}{R_{\text{yellow}}} &= \frac{F_s C_s F_w F_p}{F_s F_w F_p} = C_s \\ \frac{R_{\text{mail-er}}}{R_{\text{yellow}}} &= 0.2173 \\ C_s &= 0.2173\end{aligned}$$

The equation in (5) shows that the process of coda-based darkening at the stem level retains light [l] 21.73% of the time. This percentage of light [l] will be passed on to the word level, where the /l/ will have the chance to darken again. The 78.27% of tokens darkened at the stem level remain dark through the derivation.

Moving on to the coda-based process of darkening at the word level, we turn to the environments of *mail-er* and *mail it*. The /l/ in *mail it* is always foot-medial and so is subject to all three rounds of foot-based darkening. For coda-based darkening, *mail it* is susceptible at the stem and word levels, but escapes darkening at phrase level, as the following word *it* is now visible and the /l/ is resyllabified into the onset. Therefore, dividing *mail it* by *mail-er* will isolate the retention of light /l/ at the word level.

(6)

$$\begin{aligned}\frac{R_{\text{mail it}}}{R_{\text{mail-er}}} &= \frac{F_s C_s F_w C_w F_p}{F_s C_s F_w F_p} = C_w \\ \frac{R_{\text{mail it}}}{R_{\text{mail-er}}} &= 0.0295 \\ C_w &= 0.0295\end{aligned}$$

To conclude the analysis of the coda-based process, consider darkening in *mail it* and *mail*. The only difference between the darkening in these two environments is that the /l/ in *mail it* escapes coda-based darkening at the phrase level, as it has been resyllabified into onset position. At the stem and word levels these /l/s are in both non-foot-initial and coda position, which allows us to isolate darkening at the phrase level:

(7)

$$\frac{R_{\text{mail}}}{R_{\text{mail it}}} = \frac{F_s F_w F_p C_s C_w C_p}{F_s F_w F_p C_s C_w} = C_p$$

$$\frac{R_{\text{mail}}}{R_{\text{mail it}}} = 0.0022$$

$$C_p = 0.0022$$

In summary, the rate of retention of light [l] in the coda at each level is:

(8)

$$C_s = 0.2173$$

$$C_w = 0.0295$$

$$C_p = 0.0022$$

Crucially, light [l] is retained at a higher rate at higher levels. The relatively high rate at the stem level suggests that coda-based darkening at that level is a relatively recent and still ongoing innovation. All of this accords well with the predictions of the life cycle of phonological processes, and in particular the Variation Corollary of the Russian Doll Theorem (3).

3.3 Foot-based darkening

The aim of this section was to demonstrate how to calculate individual frequencies of application of a variable phonological process at three cyclic levels, as shown with the process of coda-based darkening. However, calculating cycle-specific rates of application of foot-based /l/-darkening proves problematic, because the data provided by Hayes data does not include the environments needed to isolate the foot-based processes at each level. There are no environments in which an /l/ begins the derivation as non-foot-initial, but advances to foot-initial position in further cycles. The possibility of relying on the /l/ in environment *free-ly* was tested, but dismissed due to the problems involved in using an /l/ which is part of a suffix, a conclusion backed up by articulatory studies showing that /l/s in suffixes may behave differently to those in stems (Lee-Kim et al. 2013).

To allow a stratal analysis, Hayes's data would have needed to include environments such as *mail-ee*, where the /l/ is foot-medial at the stem level, but becomes foot-initial at the word and phrase levels. Following

this, we would need phrases such as *mail Ann*, where the /l/ is foot-medial in the stem and word levels, but not at the phrase level. Additional data for these two environments would make it possible to isolate rates at all three separate levels.

(9)

$$R_{\text{mail-ee}} = F_s C_s$$

(10)

$$R_{\text{mail Ann}} = F_s F_w C_s C_w$$

As the environments in (9) and (10) are not available in the current data set, we have no option here but to adopt the Guyan model and take the cube root. This is not what was originally intended: the aim, again, was to calculate cycle-specific rates. However, given the limitations of Hayes's data, this is the best option for our current purposes. Therefore, hereafter we will assume that 8.5% of /l/s will darken at each level, as per (11).

(11)

$$F_{s/w/p} = \sqrt[3]{0.76691} = 0.9153$$

3.4 Frequencies for the GLA

The figures for darkening at each level can now be used to calculate the predicted frequencies for the GLA. Fig. 3 below demonstrates how this is done, using a base of 1,000,000 tokens of /l/ as an example. The idea is that, these 1,000,000 tokens pass through the stem level, a percentage are darkened, and the remaining light tokens pass onto the word level, where again they are susceptible to darkening. Fig. 3 shows the environment *yellow*, where the Guyan exponential approach is used (as explained in section 3.3) and darkening is the same at each level. The first level of branching in the trees illustrates the mapping from the underlying representation to the stem-level output. 8.5% of /l/s darken at this point, and the remaining 91.5% remain light. These relative proportions are then

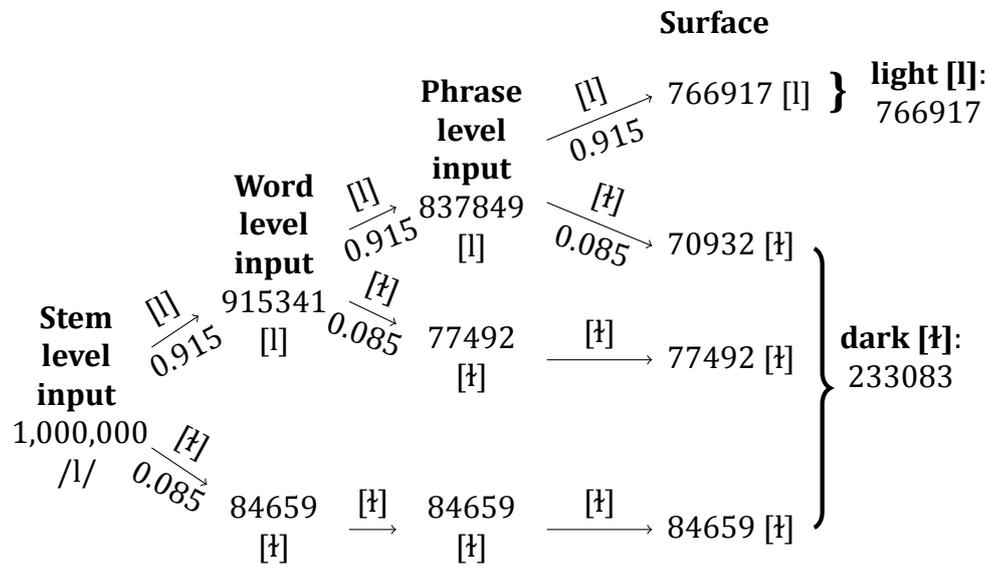


Figure 3: Tree demonstrating foot-based darkening in environment *yellow* (note that some numbers are presented as rounded but calculations are based on actual numbers).

carried over as input to the word level. Here, the remaining 91.5% of light [l] become vulnerable to darkening again.

Fig. 4 shows the derivation in a phrasal environment such as *mail*. This tree is more complex than the tree in fig. 3 as it shows the interaction of both coda and foot-based darkening processes. In the first cycle, the possibility of /l/ remaining light is calculated from two independent rules applying in either order within the same cycle: the foot-based process (at 0.915) and the coda-based stem-level process (at 0.217). The chance of darkening is 1 minus this probability.

3.5 Constraints

In order to best compare the new model with Boersma & Hayes (2001), ideally the same constraint set, shown in (12), would be employed.

As we saw in section 2.2, Hayes's approach relies on OOC constraints to account for the opaque occurrence of prevocalic dark [ɫ] before stem-suffix and word boundaries. However, OOC constraints are redundant in a stratal analysis, where opacity is accounted for by faithfulness between cycles. Analysing the same data with Stratal OT therefore means that the OOC constraints will be replaced by the faithfulness constraint IDENT[ɫ].

(12) **Hayes's Constraints**

DARK [ɫ] IS POSTVOCALIC
 *PRETONIC /l/ IS LIGHT
 PREVOCALIC /l/ IS LIGHT
 /l/ IS DARK
 IDENT-OO(Phrasal)
 IDENT-OO(Morphological)

- (13) **IDENT[ɫ]**: If an alveolar lateral is dark in the input, then its output correspondent must be dark.

IDENT[ɫ] is necessary as it accounts for the fact that /l/s which are darkened in a particular cycle remain dark in further cycles, as shown by the high frequency onset position dark [ɫ]s in *heal-ing* and *heal it*. However, the sole addition of IDENT[ɫ] to Hayes's constraint set causes problems when attempting to derive low-frequency light [l] in word-final position. The GLA generates huge negative figures for some constraints when trying to model a non-zero frequency for light [l] in environments such as *mail*. This is because the light output form [meil] is harmonically bounded by the dark output [meɫ], as shown in table 8 below.

Clearly, adding just IDENT[ɫ] to the remainder of Hayes's constraint set is not enough. The problem stems from context-free /l/ IS DARK, which makes it impossible to derive low frequency light [l] in *mail*. The effects of /l/ IS DARK are counterbalanced in Hayes's analysis by three other constraints favouring light [l], but crucially none of these favour [l] in coda position. This means that there is no way of deriving anything other than zero for output form [meil] using this constraint set. Hayes (2000:105) derives low frequency *mai*[l] by means of OOC with *mai*[l]-*er* and *mai*[l] *it*, where [l] is favoured by PREVOCALIC /l/ IS LIGHT. This is not an option in a stratal framework and not just because Stratal OT does not have OOC. More fundamentally, this is a violation of Base Priority, the idea that the base does not behave opaquely in order to correspond with a derived counterpart (Benua 1997b:239).

Moreover, although Boersma & Hayes only need to match a frequency of 0.0011% for *mail*, a cyclic model will have much higher rates of light [l] in more embedded domains. For example, at the stem level the frequency of light [l] for environment *mail* is actually 19.89%, and so a generated frequency of 0% is far from satisfactory.

The only solution, therefore, is to change the constraint set. Although it would have been preferable to use the same set of constraints as Hayes, thus keeping the experiment maximally controlled, this has nevertheless

| Input | /l/ IS DARK | /l/ IS POSTVOCALIC | PREVOCALIC /l/ IS LIGHT | PRETONIC /l/ IS LIGHT | IDENT [ɹ] |
|-------|-----------------|--------------------|-------------------------|-----------------------|-----------|
| meɪ | a. meɪ | | | | |
| | b. meɪ | * | | | |
| meɪ | a. meɪ | | | | |
| | b. meɪ | * | | | * |

Table 8: Stem-level tableau using Hayes's constraints showing a harmonically bounded [meɪ]

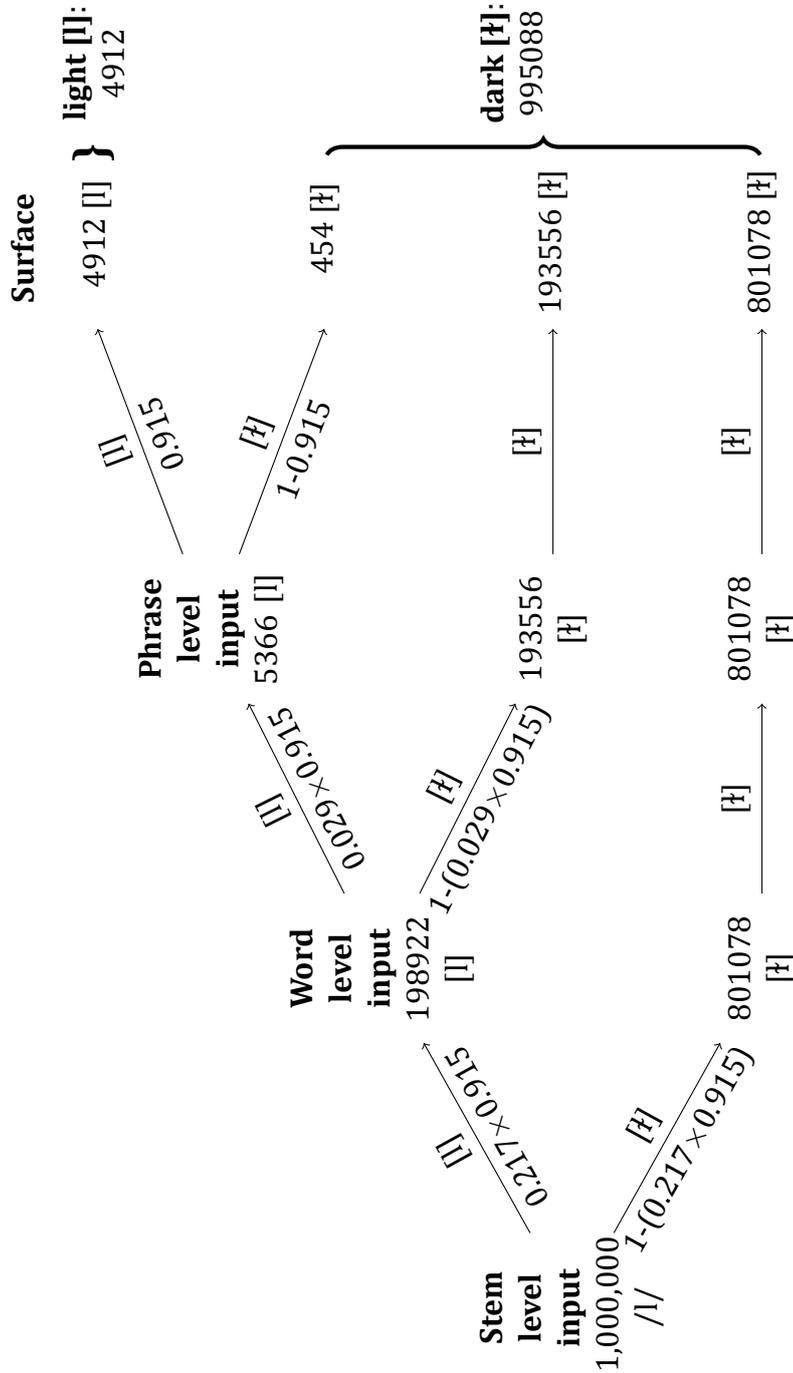


Figure 4: Tree demonstrating foot and coda-based darkening in environment *mail it*. Note that phrase level darkening only exhibits the foot-based kind, as the /l/ is resyllabified into the onset. Numbers and rates are presented as rounded, but calculations were conducted on exact figures.

proven impossible. Meanwhile, this exercise makes an important point in itself, in that the GLA's results depend on the constraint set. The new constraint set in (14) includes a faithfulness constraint for light [l], and the context-free default is [l] rather than [ɫ].

(14)

New Constraints

| | |
|-----------------------------|--|
| *[l]CODA | Coda /l/ must not be light |
| *[ɫ] | /l/ must not be dark |
| *[l] NON- Σ -INITIAL | No /l/ outside foot-initial position |
| IDENT[l] | Light /l/ in the input must remain light in the output |
| IDENT[ɫ] | Dark /l/ in the input must remain dark in the output |

This loss of consistency with the previous analysis is more than compensated for by the advantages of the cyclic approach, namely the avoidance of the problems associated with OOC (such as the *[hi:ɫɪŋ, hi:lit] dialect) and the ability to interpret level-specific rates of application in terms of the life cycle of phonological processes.

Note, moreover, that the input to the stem level is subject to Richness of the Base, a feature of OT whereby underlying representations cannot be subject to systematic restrictions. Richness of the Base means that the set of possible inputs to the grammar is universal (Prince & Smolensky 1993: 191) and therefore dark [ɫ] must be included as an input form. For our present purposes, this means that the input to the stem level may contain tokens of /l/ or /ɫ/ in any proportion. In turn, this entails that the stem-level constraint hierarchy must be able to derive the appropriate ratio of light and dark /l/ in the stem-level output regardless of what is posited in the input.

Using the procedure illustrated in figs. 3 to 4, the relative frequency of light and dark /l/ at each phonological level was calculated for five phonological environments: *light*, *yellow*, *mail-er*, *mail it* and *mail*.⁵ These frequencies were then submitted to the GLA, which acquired a stochastic OT

⁵ The inclusion of *light* within the input files is necessary. *Light* is needed to demonstrate that the grammar derives light /l/ in foot-initial onsets almost obligatorily. Boersma & Hayes's figures for light [l] in *light* are very high, at 99.956% but not quite 100%. As initial /l/s are not targeted by either of the darkening processes discussed thus far, we need an additional process to account for the fact that it is not 100%. For environment *light*, an exceptional process of darkening will be simulated, which is simply the cube root of the retention, as shown in (i). This will then be consistent with the

ranking for each of the three levels. The following section reports the results of these tests.

4 Results and discussion

The details of the training schedule and initial rankings of the GLA are shown in table 9. The learning schedule was kept the same for all three levels of the cycle. Markedness was set higher than faithfulness, which is in line with the Subset Principle, i.e. the premise that children initially opt for the most restrictive grammar possible.

| | |
|-------------------------------|---------|
| Initial ranking values | |
| <i>markedness</i> | 10,000 |
| <i>faithfulness</i> | 1,000 |
| Learning trials | 500,000 |
| Initial plasticity | 1 |
| Final plasticity | 0 |
| Testing cycles | 20,000 |

Table 9: Training schedule

The grammar was run over 500,000 trials, the results of which are presented in this section. Hereafter, *projected frequencies* refers to the results of the calculations described in section 3.4, and *generated frequencies* refers to the grammar acquired by the GLA.

Table 10 shows the ranking values acquired by the GLA at the stem level. As the underlying representation is subject to Richness of the Base (see section 3.5), the stem-level grammar must be able to derive the correct ratio of light and dark /l/ in the stem-level output, irrespective of whether the input /l/ is light or dark. Consequently, it is not surprising that the IDENT constraints are ranked low, as it is more important to obtain the correct ratios of light and dark /l/s than to preserve the input at this level.

Table 11 shows how accurately the GLA managed to match the projected frequencies at the stem level. The three final environments all have

other calculations.

$$\begin{aligned}
 \text{(i)} \quad R_{\text{light}} &= \sqrt[3]{0.99956} \\
 &= 0.99985
 \end{aligned}$$

| Constraint | Ranking Value |
|-----------------------------|---------------|
| *[l] CODA | 9,992 |
| *[ʔ] | 9,989 |
| *[l] NON- Σ -INITIAL | 9,982 |
| IDENT[ʔ] | 1,067 |
| IDENT[l] | 1,056 |

Table 10: Constraint ranking values at the stem level

| Input to level | Light [l] in output of level | | | |
|-------------------|------------------------------|------------------|------------------------------|------------------|
| | % | | cumulative total per million | |
| (Rich Base) | <i>projected</i> | <i>generated</i> | <i>projected</i> | <i>generated</i> |
| <i>light</i> | 99.99 | 100 | 999,853 | 1,000,000 |
| <i>ʔight</i> | | | | |
| <i>yellow</i> | 91.53 | 99.33 | 915,341 | 993,250 |
| <i>Yeʔow</i> | | | | |
| <i>mail-er</i> | 19.89 | 14.07 | 198,922 | 140,700 |
| <i>maiʔ-er</i> | | | | |
| <i>mail it</i> | 19.89 | 14.07 | 198,922 | 140,700 |
| <i>maiʔ it</i> | | | | |
| <i>mail</i> | 19.89 | 14.07 | 198,922 | 140,700 |
| <i>maiʔ</i> | | | | |

Average error per candidate: **0.041**

Table 11: Stem level results

the same projected frequency, as only the stem *mail* is visible at the this particular level. The average error per candidate is very low, and note that this includes the inputs with dark /ʔ/, which have the same results as their light /l/ counterparts. However, the GLA does not manage to match the frequencies derived from the data exactly. It makes the /l/ in *light* categorically light. This may not be the exact result, but the error is tiny and in the expected direction. In fact, attempting to match the conjectured frequencies of inherently scalar judgement data in this way may be a little over the top. It would arguably be acceptable to just aim for 100% light [l] here.

Turning to the word and phrase levels, we find a different situation in the ranking of constraints. As stated above, the GLA's results need to reflect the fact that a dark variant in the input remains dark in the output.

Accordingly, the faithfulness constraints must be ranked so that all dark /l/s in the input are protected; hence IDENT [ɫ] becomes top and second ranked, as shown in tables 12 and 13. This high ranking of IDENT [ɫ] captures the fact that dark [ɫ]s which are resyllabified into onset position in further cycles are opaquely preserved.

| Constraint | Ranking Value |
|-----------------------------|---------------|
| IDENT[ɫ] | 10,008 |
| *[l] CODA | 9,962 |
| IDENT[l] | 9,952 |
| *[l] NON- Σ -INITIAL | 9,948 |
| *[ɫ] | 9,944 |

Table 12: Constraint ranking values at the word level

| Constraint | Ranking Value |
|-----------------------------|---------------|
| *[l] CODA | 10,007 |
| IDENT[ɫ] | 10,006 |
| IDENT[l] | 10,000 |
| *[l] NON- Σ -INITIAL | 9,995 |
| *[ɫ] | 9,994 |

Table 13: Constraint ranking values at the phrase level

| Input to level | Light [l] in output of level | | | |
|----------------|------------------------------|------------------|------------------------------|------------------|
| | % | | cumulative total per million | |
| | <i>projected</i> | <i>generated</i> | <i>projected</i> | <i>generated</i> |
| <i>light</i> | 99.97 | 100 | 999,705 | 1,000,000 |
| <i>yellow</i> | 91.53 | 91.91 | 837,849 | 912,846 |
| <i>mail-er</i> | 91.53 | 91.91 | 182,082 | 129,310 |
| <i>mail it</i> | 2.73 | 0.01 | 5,422 | 14 |
| <i>mail</i> | 2.73 | 0.01 | 5,422 | 14 |

Average error per candidate: **0.015**

Table 14: Word level results

The generated results at the word level successfully manage to match the projected results more accurately when considering the percentages

at this level alone. However, the errors incurred in the previous cycle are apparent, as shown by the cumulative total columns. This shows that a cyclic approach needs to work within a tight degree of accuracy, as small deviations in higher levels will be amplified at lower levels.

Note that the figures for input dark /t/ are omitted in tables 14 and 15. This is because at the word and phrase levels all the dark inputs simply have to remain dark. The GLA learns this easily, matching the projected frequency with 100% accuracy. This also brings down the average error per candidate, as shown on the bottom right of each table.

| Input to level | Light [l] in output of level | | | | |
|-------------------|------------------------------|------------------|------------------------------|------------------|---------|
| | % | | cumulative total per million | | |
| | <i>projected</i> | <i>generated</i> | <i>projected</i> | <i>generated</i> | % error |
| <i>light</i> | 99.99 | 100 | 999,559 | 1,000,000 | 0.04 |
| <i>yellow</i> | 91.53 | 96.50 | 766,917 | 880,896 | 11.40 |
| <i>mail-er</i> | 91.53 | 96.50 | 166,667 | 124,784 | 4.19 |
| <i>mail it</i> | 91.53 | 96.50 | 4,963 | 14 | 0.49 |
| <i>mail</i> | 0.20 | 0.006 | 11 | 0 | 0.0001 |

Average error per candidate: **0.012**

Table 15: Phrase level results

The phrase level results in table 15 show that, for most environments, the disparity between the projected frequencies and those generated by the GLA is very small. In the environments *light*, *mail it* and *mail* the error is always less than half a per cent, with *mail-er* a little higher at just over 4%. However, the environment *yellow* fails to generate an acceptable margin of error, at a much larger 11.40%. This is most likely due to the fact that this environment is precisely the one where Hayes's data made it impossible to run cycle-specific calculations. Darkening in this environment is entirely foot-based (as the /l/ is not in the coda at any level), and $F_{s/w/p}$ was calculated by taking the cube root in the manner of Guy, making the figure less reliable.

In fact, there are good reasons to believe that the expedient adopted in 11 above may not be an accurate representation of the facts. If a foot-based process represents a relatively advanced stage of darkening, as explained in section 2.4, it would be reasonable to predict that it has not yet reached the level of the stem, or perhaps even the word. To improve the analysis, one might consider comparing the scenario above with alternatives where foot-based darkening has not progressed beyond the word

level, or even the phrase level. If either of these analyses can avoid the high discrepancy in environment *yellow* without increasing error rates in other environments, then we may conclude that foot-based /l/-darkening has not reached the stem level in this dialect at this moment in time. However, data for phonological environments such as *mail-ee* and *mail Ann* would be a further improvement, removing the need for a comparison of hypothetical scenarios and making it possible to calculate accurate cycle-specific rates.

5 Conclusion

This paper has illustrated how SSOT can model the variable, morphosyntactically conditioned application of /l/-darkening in present-day American English, as reflected in the frequencies provided by Boersma & Hayes (2001), with minimal error. It has been shown that an SSOT approach preserves the advantages of Stochastic OT, in that it is able to model variation, and the advantages of Stratal OT, in that it captures the role of morphosyntactic factors without stipulating innate rankings.

Moreover, I have demonstrated an innovative way of working out the rate of application of a phonological process at each cyclic level. This method, when applied to the /l/-darkening data, corroborates the predictions of the Variation Corollary of the Russian Doll Theorem, stated in (3). In American English, coda-based darkening has run to completion at the phrase and word levels, applying approximately 97% of the time in both strata. At the stem level, in contrast, its rate of application remains significantly lower: around 78%. This reflects the diachronic facts: stem-level darkening is a relatively recent innovation.

Refining this model of /l/-darkening further would require several adjustments. First, one would need empirical data from more environments, crucially including expressions like *mail-ee* and *mail Ann*. Such data would allow one to calculate cycle-specific rates of foot-based darkening, avoiding the expedient of assuming equal rates of foot-based darkening at all levels, in the manner of Guy. Secondly, the use of actual phonetic data, rather than frequencies inferred from well-formedness judgements, would undeniably give a greater degree of reliability.

In addition, this paper does not claim to provide a full simulation of learning. That is, I have not automated the whole procedure for going from observed surface frequencies in different environments to the full SSOT grammar. This remains a goal for further research. What has been demonstrated, however, is the existence of an SSOT grammar that performs as well as Boersma & Hayes's parallel account, but without the ty-

phonological shortcomings of the latter. SSOT retains the ability to model phonological variation in the context of a more satisfactory theory of the morphosyntax-phonology interface, which, moreover, makes direct links with the life cycle of phonological processes.

It should also be pointed out that many phonetic studies of /l/-darkening have found little intra-speaker variability (Turton 2014a). That raises the issue of whether Hayes's data reflect variation at the level of the speaker or the level of the speech community. It may even reflect variation at the level of the English-speaking world. This is the nature of the judgement experiment carried out: we have no way of knowing. Until articulatory data becomes more efficient to measure, investigating intra and inter-speaker variability is probably a job best left to a carefully controlled acoustic analysis.

In sum, English /l/-darkening poses formidable challenges to phonological theory. This paper has tackled one of these challenges, namely, how to model variable morphosyntactically conditioned opacity. Such morphosyntactic sensitivity is a reflection of its historical trajectory, as phonological processes become more integrated with structure over time. The synchronic grammar, as analysed here, is merely an insight into an ever-changing system. The conclusion presented here is that modelling morphosyntactically conditioned variable phonological processes is best tackled with a stratal approach paying due consideration to the diachronic evolution of phonological processes.

Comments invited

PiHP relies on post-publication review of the papers that it publishes. If you have any comments on this piece, please add them to its comments site. You are encouraged to consult this site after reading the paper, as there may be comments from other readers there, and replies from the author. This paper's site is here:

<http://dx.doi.org/10.2218/pihph.1.2016.1697>

Acknowledgements

I would like to thank Ricardo Bermúdez-Otero for extensive comments and guidance, as well as Michael Ramsammy, Yuni Kim and Bruce Hayes for additional advice and suggestions. I would like to acknowledge audiences at The University of Pennsylvania, the Manchester Phonology Meeting and the PhilSoc Symposium at Oxford for their questions and com-

ments. This work was funded by the Arts and Humanities Research Council.

Author's contact details

Danielle Turton
 School of English Literature, Language and Linguistics
 Newcastle University
 Newcastle NE1 7RU
 United Kingdom
danielle.turton@ncl.ac.uk

References

- Anttila, Arto. 1997. Deriving variation from grammar. In Frans Hinskens, Roeland van Hout & W. Leo Wetzels (eds.), *Variation, change and phonological theory*, 35–68. Amsterdam and Philadelphia, PA: John Benjamins.
- Anttila, Arto. 2007. Variation and optionality. In Paul de Lacy (ed.), *The Cambridge handbook of phonology*, Cambridge: Cambridge University Press.
- Baković, E., Eric. 2001. Nasal place neutralization in Spanish. Tech. rep. University of Pennsylvania Working Papers in Linguistics.
- Benua, Laura. 1995. Identity effects in morphological truncation. In Jill Beckman, Suzanne Urbanczyk & Laura Walsh Dickey (eds.), *Papers in Optimality Theory* University of Massachusetts Occasional Papers in Linguistics 18, 77–136. Amherst, MA: GLSA, University of Massachusetts.
- Benua, Laura. 1997a. Affixes classes are defined by faithfulness. *University of Maryland Working Papers in Linguistics 5. Selected Phonology Papers from Hopkins Optimality Theory Workshop 1997 / University of Maryland Mayfest 1997*, 1–27.
- Benua, Laura. 1997b. *Transderivational identity: phonological relations between words*. Amherst: University of Massachusetts Doctoral dissertation. Available at ROA-259-0498, Rutgers Optimality Archive, <http://roa.rutgers.edu>.
- Bermúdez-Otero, Ricardo. 1999. *Constraint interaction in language change: quantity in English and Germanic [opacity and globality in phonological change]*.: University of Manchester and Universidad de Santiago de Compostela Doctoral dissertation.

- Bermúdez-Otero, Ricardo. 2007. Word-final prevocalic consonants in English: representation vs derivation. Paper presented at the Old World Conference in Phonology 4, Rhodes. Handout available at <http://www.bermudez-otero.com/OCP4.pdf>.
- Bermúdez-Otero, Ricardo. 2010. The history of English phrase-level syllabification: A sequel to Minkova (2003: ch. 4). Paper given at the inaugural event of the Historical Linguistics Research Cluster, Cambridge, 10 May 2010.
- Bermúdez-Otero, Ricardo. 2011. Cyclicity. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, vol. 4, 2019–2048. Malden, MA: Wiley-Blackwell.
- Bermúdez-Otero, Ricardo. 2015. Amphichronic explanation and the life cycle of phonological processes. In Patrick Honeybone & Joseph C. Salmons (eds.), *The Oxford handbook of historical phonology*, 374–399. Oxford: Oxford University Press.
- Bermúdez-Otero, Ricardo. Forthcoming. Stratal phonology. In Anna R. K. Bosch Hannahs, S.J. (ed.), *The Routledge handbook of phonological theory*, Abingdon: Routledge.
- Bermúdez-Otero, Ricardo & Ana R. Luís. 2009. Cyclic domains and prosodic spans in the phonology of European Portuguese functional morphs. Workshop on the Division of Labour between Morphology and Phonology & Fourth Meeting of the Network 'Core Mechanisms of Exponence', Meertens Instituut, Amsterdam.
- Bermúdez-Otero, Ricardo & Graeme Trousdale. 2012. Cycles and continua: on unidirectionality and gradualness in language change. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 691–720. Oxford: Oxford University Press.
- Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21, 43–58.
- Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32(1), 45–86.
- Carter, Paul. 2003. Extrinsic phonetic interpretation: spectral variation in English liquids. In Richard Ogden John Local & Rosalind Temple (eds.), *Phonetic interpretation: Papers in laboratory phonology vi*, 237–252. Cambridge: Cambridge University Press.
- Carter, Paul & John Local. 2007. F2 variation in Newcastle and Leeds English liquid systems. *Journal of the International Phonetic Association* 37(2), 183–199.

- Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
- Coetzee, Andries. 2004. *What it means to be a loser: non-optimal candidates in optimality theory*: University of Massachusetts, Amherst Doctoral dissertation.
- Coetzee, Andries W. & Joe Pater. 2011. The place of variation in phonological theory. In John Goldsmith, Jason Riggle & Alan Yu (eds.), *The handbook of phonological theory*, 401–434. Oxford: Blackwell 2nd edn.
- Gick, Bryan, Fiona Campbell, Sunyoung Oh & Linda Tamburri-Watt. 2006. Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics* 34(1), 49 – 72.
- Giegerich, H. 1992. *English phonology: An introduction*. Cambridge: Cambridge University Press.
- Giles, Stephen B. & Kenneth L. Moll. 1975. Cinefluorographic study of selected allophones of English /l/. *Phonetica* 31, 206–227.
- Guy, Gregory R. 1991a. Contextual conditioning in variable Lexical Phonology. *Language Variation and Change* 3(2), 223–239.
- Guy, Gregory R. 1991b. Explanation in variable phonology: an exponential model of morphological constraints. *Language Variation and Change* 3(1), 1–22.
- Halle, Morris & K. P. Mohanan. 1985. Segmental phonology of Modern English. *Linguistic Inquiry* 16, 57–116.
- Harris, John. 1989. Towards a lexical analysis of sound change in progress. *Journal of Linguistics* 25(1), 35–56.
- Harris, John. 2003. Release the captive coda: the foot as a domain of phonetic interpretation. In John Local, Richard Ogden & Rosalind Temple (eds.), *Phonetic interpretation: papers in laboratory phonology vi*, 103–129. Cambridge: Cambridge University Press.
- Harris, John. 2006. Wide-domain r-effects in English. *UCL Working Papers in Linguistics* 18.
- Harris, John & Jonathan Kaye. 1990. A tale of two cities: London glottalling and New York City tapping. *The Linguistic Review* 7(3), 251–274.
- Hawkins, S. 2004. Influence of syllable-coda voicing on the acoustic properties of syllable-onset /l/ in English. *Journal of Phonetics* 32(2), 199 – 231.
- Hayes, Bruce. 2000. Gradient well-formedness in Optimality Theory. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.), *Optimality Theory: phonology, syntax, and acquisition*, 88–120. Oxford: Oxford University Press.
- Huffman, Marie K. 1997. Phonetic variation in intervocalic onset /l/s in English. *Journal of Phonetics* 25(2), 115 – 141.

- Jensen, John T. 1993. *English phonology*. Amsterdam: John Benjamins.
- Jensen, John T. 2000. Against ambisyllabicity. *Phonology* 17, 187–235.
- Kahn, Daniel. 1976. *Syllable-based generalizations in English phonology*: MIT Doctoral dissertation.
- Kenstowicz, Michael. 1996. Base-identity and uniform exponence: alternatives to cyclicity. In Jacques Durand & Bernard Laks (eds.), *Currents trends in phonology: models and methods*, vol. 1, 363–393. Salford: European Studies Research Institute, University of Salford.
- Kiparsky, P. 1998. Paradigm effects and opacity. *Ms., Stanford University*.
- Kiparsky, Paul. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10(3), 421–441.
- Kiparsky, Paul. 1982. Lexical morphology and phonology. In In-Seok Yang for the Linguistic Society of Korea (ed.), *Linguistics in the morning calm: selected papers from sicol-1981*, vol. 1, 3–91. Seoul: Hanshin Publishing Company.
- Kiparsky, Paul. 2000. Analogy as optimization: ‘exceptions’ to Sievers’ Law in Gothic. In Aditi Lahiri (ed.), *Analogy, levelling, markedness: principles of change in phonology and morphology*, 15–46. Berlin: Mouton de Gruyter.
- Labov, William. 1969. The logic of non-standard English. *Florida FL Rep.*
- Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world’s languages*. Oxford: Blackwell.
- Lee-Kim, Sang-Im, Lisa Davidson & Sangjin Hwang. 2013. Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology* 4(2), 475–511.
- Lin, Susan. 2011. *Production and perception of prosodically varying inter-gestural timing in American English laterals*: University of Michigan dissertation.
- McMahon, April M. S. 2000. *Lexical Phonology and the history of English*. Cambridge: Cambridge University Press.
- Minkova, Donka. 2003. *Alliteration and sound change in early English* Cambridge Studies in Linguistics 101. Cambridge: Cambridge University Press.
- Myers, Scott. 2000. Boundary disputes: the distinction between phonetic and phonological sound patterns. In Noel Burton-Roberts, Philip Carr & Gerard Docherty (eds.), *Phonological knowledge: conceptual and empirical issues*, 245–272. Oxford: Oxford University Press.
- Olive, Joseph P., Alice Greenwood & John Coleman. 1993. *Acoustics of American English speech: a dynamic approach*. New York: Springer Verlag.

- Prince, Alan & Paul Smolensky. 1993. Optimality Theory: constraint interaction in generative grammar. Tech. Rep. RuCCS-TR-2 Rutgers University Center for Cognitive Science.
- Ramsammy, Michael. 2011. *The realisation of coda nasals in Spanish*: University of Manchester dissertation.
- Ramsammy, Michael. 2015. The life cycle of phonological processes: accounting for dialectal microtypologies. *Language and Linguistics Compass* 9(1), 33–54.
- Recasens, Daniel & Aina Espinosa. 2005. Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /l/: evidence from two Catalan dialects. *Journal of the International Phonetic Association* 35, 1–25.
- Roach, Peter, Daniel Jones & James Hartman. 2006. *Cambridge English pronouncing dictionary*. Cambridge University Press, 16th edn.
- Sproat, Richard & Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of phonetics* 21(3), 291–311.
- Turton, Danielle. 2014a. *Variation in English /l/: Synchronic reflections of the life cycle of phonological processes*: University of Manchester dissertation.
- Turton, Danielle. 2014b. Some /l/s are darker than others: accounting for variation in English /l/ with ultrasound tongue imaging. *Penn Working Papers in Linguistics* 20(2).
- Turton, Danielle. Forthcoming. Categorical or gradient? An ultrasound investigation of /l/-darkening and vocalisation in varieties of English. *Laboratory Phonology*.
- Vennemann, Theo. 1972. Phonetic analogy and conceptual analogy. In Theo Vennemann & Terence H. Wilbur (eds.), *Schuchardt, the Neogrammarians, and the transformational theory of phonological change: four essays*, 181–204. Frankfurt am Main: Athenäum Verlag.
- Yuan, Jiahong & Mark Liberman. 2009. Investigating /l/ variation in English through forced alignment. In *INTERSPEECH 2009: 10th Annual Conference of the International Speech Communication Association, Brighton, UK, 6-10 September 2009*, 2215–2218.
- Yuan, Jiahong & Mark Liberman. 2011. /l/ variation in American English: A corpus approach. *Journal of Speech Sciences* 1(2), 35–46.