Rethinking English vowel alternation

Jonathan B. Alcantara

This paper addresses the classic problem of vowel alternation in the English lexicon. It argues that, at least for the Latinate portion of the grammar, this alternation might be better regarded as determined by a constraint-based lengthening process rather than shortening, as is traditionally assumed.

1. Introduction

In English, there is a class of vowel alternations usually characterized as indicating an underlying phonemic difference in quantity (cf. Halle 1977; Myers 1987; Kiparsky 1982a, among others). The following diagram, adapted from Myers (1987: 485), illustrates this type of alternation:

(1)  [ay] - [i]:  divine / divinity  /i/-/i/
     [iy] - [e]:  deep / depth    /ɛ/-/e/
     [ey] - [æ]:  nation / national /a/-/a/
     [ow] - [a]:  cone / conic    /o/-/o/

Chomsky and Halle (1968) used three shortening rules, discussed below, to account for these alternations. More recent work in prosodic and metrical phonology has attempted to apply the prosodic categories used so persuasively to account for stress assignment and other processes, both phonological and syntactic (e.g. Nespor and Vogel 1983; Zec and Inkelas 1990; Hayes 1993) to this problem. However, no analysis has managed to satisfactorily account for both vowel alternation and stress assignment. Here it will be argued that a prosodic account can indeed be fruitful in this area, if we abandon certain assumptions concerning underlying representations, some of which originated in the pre-generative era.

In this paper, I will first discuss the forms involved in this vowel alternation, and the principal theories and solutions which have been offered for this problem. I will then point out the reasons why these solutions are inadequate; as evidence I offer both my own observations and the ideas of others who have noticed similar problems with these

---

1 The relationships of [uw] : [ʌ], [aw] : [ɔː] and [ɔy] : [ɔː], e.g. deduce / deduction, profound / profundity, boil / ebullition, are not as clear, but additionally do not form a significant part of the alternations in question. (cf. Chomsky and Halle 1968; Halle 1977).
accounts of vowel alternation. Subsequently, previously unnoticed facts about the distribution of vowel length in the surface forms will be pointed out, and that information, coupled with recent ideas concerning the interaction of prosodic constituents and structural constraints, will suggest a novel analysis based upon a revision of the underlying lexical forms for these words. Finally, the new analysis will be shown to help remove some troublesome complexities from the grammar, the need for them now being eliminated.

2. Background to the problem
2.1 The data: vowel alternation within derivational paradigms

Myers (1987) is useful as a starting point for a description of the problem, as he nicely summarizes the environments for vowel alternation. Myers identifies the following types of vowel alternation found in morphologically related forms. The sample pairs given below are taken from his much longer lists. The first three sets involve regular alternations based on suffix-type:

(2a) Consonant-initial suffixes (Myers 1987: 488-9):
   (/-t/, /-th/, /-tive/, /-tion/, /-ture/)
   
   keep / kept          wide / width
   describe / descriptive         convene / convention
   discreet / discretion         revise / revision
   scrible / scripture           five / fifth

(2b) Multisyllabic vowel-initial suffixes (Myers 1987: 494-5):
   (/-ity/,/-ative/,/-itive/,/-atory/,/-ible/,/-able/,/-ify/,/-ite/,/-ual/,/-acy/ )

       sincere / sincerity    sane / sanity
       derive / derivative    appeal / appellate
       prime / primitive      define / definitive
       exclaim / exclamatory  oblige / obligatory
       creed / credible       placate / placable
       mode / modify          vile / vilify
       sole / solitude        grateful / gratitude
       grade / gradual        rite / ritual
       supreme / supremacy    conspire / conspiracy
(2c) Monosyllabic vowel initial suffixes (Myers 1987: 499-500):
( /-ic/, /-id/, /-ish/, /-ule/)

cone / conic  state / static
pale / pallid  vapor / vapid
final / finish  Spain / Spanish
globe / globule  node / nodule

In all these cases, addition of a suffix to a "base" (unsuffixed) form showing a long vowel apparently causes the vowel to shorten. These suffixes are, for the most part, those described in Lexical Phonological work as "level I", that is, those suffixes which also affect stress (e.g., Kiparsky 1982a,b; Halle and Mohanan 1985). These may be contrasted with level II affixes such as /-ship, -ness, -ing/, which are generally understood as not altering the location of main stress.

The shortening environments described above are familiar from Chomsky and Halle (1968), who proposed three separate shortening (or "laxing") rules to cover these cases, namely CC Shortening for set (2a), -ic/-id/-ish Shortening for set (2c), and Trisyllabic Shortening for set (2b).

Additionally, Myers notes the shortening seen in Latinate prefixes, in suffixed words of the following kind (p. 500):

(3) reféute / réfutation  déforma / déformation
préäsent / présentation  provide / provident

In these words, addition of a level I suffix to the base form causes a long vowel in the prefix to shorten; furthermore, the main stress of the word shifts off the base to the prefixal syllable.

The preceding cases may be compared with a series of vowel-initial suffixes of level I which do not produce such vocalic alternations; the suffixed forms contain the same vowel as the base forms:
Suffixes producing no length alternations:

( /-al/, /-ort/, /-ory/, /-ary/, /-ous/)  

scribe / scribal                 tone / tonal
advise / advisor                 propel/ propellor
advise / advisory                supervise / supervisory
fame / famous                    desire / desirous

2.2 The account of Myers (1987)

According to Myers (1987: 486), "vowel quantity in English depends to a great extent on stress." The earliest metrical accounts of English stress, such as Liberman and Prince (1977), Hayes (1982), and Selkirk (1980), moved toward doing away with the segment-based stress rules of Chomsky and Halle (1968), which contained variables, disjunctive ordering, and the feature [± stress], replacing them with prosodic or metrical analyses that utilized hierarchically arranged tree structures or grids to derive heretofore representationally complex prominence relations. These representations, dependent only upon local interactions, utilize their inherent cumulative properties to model the apparently long-distance relations seen in, for example, secondary stress assignment and phrasal stress.

However, the early metrical accounts conspicuously did not address the issue of vowel alternation, which also required non-local, stress-dependent rules in Chomsky and Halle (1968). For example, the influential account of stress in Hayes (1982) does not treat any cases involving vowel alternation. Since Chomsky and Halle's vowel alternation rules were dependent upon their stress rules, improving upon their stress rules invites similar improvement for the vowel alternation rules.

If vowel length is indeed dependent upon or related to stress, the same mechanisms used to represent stress should also be pertinent for explaining vowel length alternations. In the relevant literature that touches upon the problem of vowel alternation (Chomsky and Halle 1968; Myers 1987; Prince 1990; Halle and Vergnaud 1987; Burzio 1993), stress is always a factor in the formulation of vowel shortening rules. Consequently, such accounts are necessarily dependent upon their theoretical framework for stress assignment in their explanation of vowel alternation.

The account of Myers (1987) has been widely cited in the lexical phonological literature in regard to this problem of vowel alternation. Myers follows Selkirk (1984), in constructing his stress placement rules. Under this approach, the only relevant prosodic category available to the analysis is the syllable. Higher-level domains are handled by a
grid-based representation which appears to govern only stress. Thus, any rules which affect segments are bound to the syllable domain.

Myers' (1987 : 488) solution proceeds from an important generalization that he noticed concerning the distribution of vowel quantity in the English lexicon: "within roots, long vowels appear only in open syllables. A nonfinal closed syllable must generally have a short vowel..." Single final consonants in English, as in tape or seem, are regarded as extrametrical, following Hayes (1982), leaving these syllables "in effect open" as well. True closed syllables in final position must therefore end in two consonants, e.g. predict. The small class of words, such as basic, deictic, danger, which unexpectedly show long vowels in the shortening environment are regarded by Myers as "idiosyncratic lexical exceptions," which he lists in an appendix to the article.²

On the basis of this generalization, Myers (1987 : 491) proposes a rule of Closed Syllable Shortening which shortens a long vowel before a tautosyllabic consonant, and notes that this simply restates, in slightly different terms, earlier solutions given by Stampe (1972) and Chomsky and Halle (1968) ("CC Shortening"). As this rule appears to represent a constraint whose effects are also seen in the lexicon itself, Myers (1987 : 511) formalizes this restriction by using a syllable template proposed for English by Borowsky (1986), wherein a well-formed syllable is of the form C* V (X), where X can be fulfilled by C or V (C* indicates an optional onset). This template allows only for one optional post-vocalic timing slot, which can be either filled by a consonant or the "second half" of the long vowel, but not both.

The consonant-initial suffixes from (2a) above, (/l-tl/, /l-th/, /livel/, /l-tionl/, /l-turel/), when added to underlyingly long "roots" with final, extrametrical consonants, produce a situation in which the root-final consonant becomes syllable-closing. No longer word-final, it cannot be extrametrical, following the peripherality constraints on extrametricality of Hayes (1982), and thus triggers the shortening rule. In terms of the syllable template, once the formerly extrametrical consonant is syllabified as part of the word-final syllable, linking to the final 'C' slot of a (C)VC template, the long vowel loses its second timing slot in the template, and thus "shortens":

² Myers excuses a large group of words, such as pint, beast, round, as ending in an extrametrical cluster of final coronals, following Kiparsky (1981). As these words for the most part do not engage in Level I suffixation, they will not be treated here.
(5) \(\text{convne : convention}\)

\[
\begin{array}{lll}
\text{CVV} & \text{CVVC} & \text{CVC} \\
\text{con. ve:<n>ex + tio<n>ex} & \rightarrow & \text{con. ve:n. tio<n>ex} & \rightarrow & \text{con. ven. tio<n>ex} \\
\text{Suffixation} & \text{Loss of Extrametricality} & \text{Shortening}
\end{array}
\]

For the vowel-initial suffix cases listed in (4), such as /-al/, /-or/, /ory/, the formerly extrametrical final consonant resyllabifies as the onset of these onsetless suffixes, and does not affect the shape of the preceding syllable. As a result, no shortening takes place:

(6) \(\text{fame : famous}\)

\[
\begin{array}{l}
\text{fa:<m>ex + ou<s>ex} \rightarrow \text{fa: mou<s>ex} \\
\text{Suffixation} & \text{Loss of Extrametricality}
\end{array}
\]

Having established a shortening environment, Myers moves on to a less transparent set of cases: those in which a vowel-initial suffix triggers shortening, such as the examples in (2b). For such cases, Chomsky and Halle (1968) proposed the rule of Trisyllabic Shortening (or Laxing); Myers instead suggests that these cases are once again instances of Closed Syllable Shortening. He notes that, in every case, "the shortened vowel is always in a stressed syllable immediately followed by an unstressed one." (Myers 1987: 495) He cites, among others, Stampe (1978), Selkirk (1982b) and Borowsky (1985) for evidence of a "stress-sensitive rule of Resyllabification", which resyllabifies sequences like V'. CV as V'C. V. Thus, a form like vilify, derived from the monosyllabic adjective vile, would undergo the following type of change:

(7) \(\text{vľ:<l> + i. fy} \rightarrow \text{vľ: li. fy} \rightarrow \text{vľ:l. i. fy} \rightarrow \text{vľ. i. fy}\)

After Resyllabification, the stressed syllable is closed, and the long vowel shortens by Closed Syllable Shortening.

The suffixes in (2b) condition the Resyllabification rule because they are all analyzed as disyllabic and unstressed, with final extrametrical syllables, following the principle of affix extrametricality (Hayes 1982). A parallel solution applies to the words with monosyllabic vowel-initial suffixes from set (2c) above ( /-ic/, /-id/, /-ish/, /-ule/), because, again according to Hayes (1982), these suffixes are idiosyncratically exempt.
from suffix extrametricality. This exemption produces an identical environment for shortening:

(8) \[ \text{stá:<t> + ic} \rightarrow \text{stá:. tic} \rightarrow \text{stát. ic} \rightarrow \text{stát. ic} \]

Suffixation Resyllabification Shortening

Although Myers does not explicitly discuss the suffixes that do not cause shortening, set (4) above (/-al/, /-or/, /-ory/, /-ary/, /-ous/), presumably these cases are not subject to Resyllabification because the entire suffix remains extrametrical throughout the affixation process\(^3\). Thus, revising (6) above, root-final consonants remain peripheral, and never lose their extrametricality:

(9) \[ \text{fa:<m> + <ous>} \rightarrow \text{fa: <mous>} \]

Finally, Myers similarly handles the cases of prefix shortening, set (3) above. To create the proper stress environment for Resyllabification, Myers invokes a rule of Sonorant Destressing (cf. Hayes 1982; Kiparsky 1979) which destresses a heavy syllable containing a sonorant when following a stressed syllable:

(10) \[ \text{de:. fé} + \text{ence} \rightarrow \text{dé:. fe. rence} \]

Suffixation Sonorant Destressing

\[ \rightarrow \text{dé:f. e. rence} \rightarrow \text{déf. e. rence} \]

Resyllabification Shortening

Thus, Myers has converted the three shortening rules of Chomsky and Halle (1968), i.e. CC Shortening, -ic/-id/-ish Shortening, and Trisyllabic Shortening, into one rule of Closed Syllable Shortening, relying on extrametricality and Resyllabification to create the proper shortening environments.

2.3 Problems with Myers' analysis

While Myers' account is consistent, there are reasons to question a number of his mechanisms and assumptions. One problem is noted by Myers himself (1987: 514-5): for roots ending in two consonants, the following situation would be reached upon Resyllabification:

\(^3\) For the treatment of /-ary/ and /-ory/ as monosyllabic suffixes, see, for example, Kager (1989).
At the stage where the "root" syllable is /fáls/, we should expect a compensatory shortening parallel to that found in long-vowel roots, due to the syllable template's failure to license CVCC, in which a consonant would be deleted, e.g. *fäsity or *falité. To remove this false prediction, Myers proposes for English a structure preservation rule which preserves timing units corresponding to separate feature matrices. Thus, consonants will not be left without a timing slot, as in the case of falsity, but long vowels may lose a timing slot without completely delinking the feature matrix of the vowel in question. As a similar deletion of consonants does occur at level II (e.g., sign, bomb, damn), Myers' constraint must apply only to level I.

To preserve a consistent treatment of the syllable template⁴, Myers regards the constraint as "undoing" Resyllabification. Since the onset is not lost in these cases, Myers assumes that it returns to its place. The rule is presumably also blocked in the case of exceptional words like notify, basic, whose final consonants are perhaps marked as not subject to Resyllabification. At this point, it appears that Resyllabification, introduced as a general syllabification rule, seems only to function in those cases where an explanation for vowel shortening is needed.

Myers cites other evidence for Resyllabification, but there are difficulties with these as well. He notes that Selkirk (1982b) claims that the flap [D] is restricted in English to syllable-final position as an allophone of /t/, and appears in ambisyllabic position "by virtue of Resyllabification. The /t/ is attracted out of the unstressed final syllable and tacked onto the end of the preceding stressed syllable, where it is transformed into the syllable-final allophone [D]." (Myers 1987: 496).

However, directly linking Resyllabification and the appearance of [D] is fraught with problems. Firstly, the assumption that [D] is the syllable-final allophone of /t/ does not stand up. For example, word-final /t/, while also syllable final, is never flapped unless it becomes the onset of a following vowel-initial word, e.g., get out [geDawt]. In many cases where word-internal /t/ is clearly syllable final, as in atlas, Atkins, flap is again never encountered. Syllable-final /t/ produces glottalized [tʔ] rather than flapped [D] (Kahn 1976). Selkirk (1982b) needed to introduce a new feature [release] to distinguish

---

⁴ Presumably, words like function and juncture, which have the illegal template *CVCC in their lexical entries, are somehow allowed to be exceptional.
between flapped and glottalized syllable-final /l/ in order to maintain her claim. If the proposal that flap is the syllable-final allophone of /l/ is dropped, following Kahn (1976), Resyllabification becomes an obstacle to flapping.

In addition, flapping extends beyond the stress environment delineated by Myers; words like cápital [kæpɪˈdæl] and legáality [liːɡæˈælɪti], for example, show flaps (in some pronunciations, including my own) in the second rather than the first stressless syllable following a stress\(^5\). Flapping is a post-lexical rule, occurring also across word boundaries, e.g., colloquial "I wanted to tell you" as [ aɪ wænərɪs tɛlɪʊə]. According to Kahn (1976: 95), flap can even appear across a word boundary before a stressed syllable, as in "ge[D] Ann home on time." In Kahn's framework, flap may appear in all unstressed syllables, not just those following a stressed syllable. Thus, Myers' environment for Resyllabification is clearly not identical to that of flapping.

Furthermore, the post-lexical nature of flapping suggests that the Resyllabification rule continues to be in effect beyond level I. But this would produce incorrectly shortened forms in the case of words like boating, floater, motor, unless the syllable template (which governs the shortening effect) simply fails to operate after level I. Similarly, Halle and Vergnaud (1986: 252-3) reject Myers' account of shortening through Resyllabification, due to the lack of any observable difference between the realization of resyllabified and non-resyllabified onsets, such as the /l/ seen in pairs like phallus and phallic.

Additionally, linking Resyllabification and flapping implies that when Resyllabification is blocked, flapping should also fail to occur. Thus, lexical exceptions to shortening like notify should not show a flap, as the idiosyncratic failure of Resyllabification to apply to the /l/ of this root, preventing shortening, leads us to expect *[nɒwəˈfæɪ] rather than the attested [nɒwˈdræɪ], with flap. Milliken (1989 : 254-5) also demonstrates that a Resyllabification rule which feeds flapping cannot occur past level one without predicting incorrect results. If flapping is indeed produced by a resyllabification process, this process has no relation to that which putatively triggers shortening, and indeed seems to occur post-lexically rather than at level I.

Myers also refers to Borowsky (1986) for evidence of Resyllabification. Borowsky (1986 : 262) claims that deletions of the following type are also the result of Resyllabification:

\(^5\) Borowsky(1986) tries to account for forms like capital by suggesting that the underlying form is /capitl/, but admits that this approach cannot deal with forms like divinity, repetitive.
(12)  

h-deletion  
vehicular / vé[h]icle  
prohibit / pro[h]ibition

y-deletion  
amnúity / ánnyupalatalization  
cónstitute / cons[tʃ]utive  
voluminous / vól[y]ume

lé

In each case, stressed syllables show a different phonological treatment than unstressed syllables following a stressed syllable. Borowsky claims that when Resyllabified, /h/ is deleted, /y/ is retained post-coronally, and coronals before such retained glides are palatalized, respectively. However, there are complications. In the first place, any deletion of /h/ or /y/ will violate Myers' structure preservation constraint, even if it is restricted to level I as noted above. Furthermore, both the palatalization environment and the retention of /y/ in the preceding syllable require an illegal (C)VCC syllable template (e.g. annual /árny. u. al/). According to Myers' explanation, the glide should be pushed back (or confined) to the onset position, where, following Borowsky, we would expect it to be deleted. Additionally, the proposed Resyllabification of /h/ fails to shorten the preceding vowel.

Finally, the rule of Resyllabification may be criticized for a more fundamental reason: it removes onsets from syllables, creating onsetless syllables which follow closed syllables. This option is specifically prohibited by the Onset Principle of Itô (1989), who also notes that such a syllabification pattern never occurs in any known language. Following this, it should be clear that Resyllabification is not a workable solution; to accept a violation of the otherwise universal Onset Principle, we would require much stronger proof than that offered by these accounts.

Myers, in a footnote (p. 496) allows for an alternative explanation of the facts of Resyllabification via a rule of ambisyllabicity, as defined in Kahn (1976). This would allow the onset in question to remain attached to its syllable, while simultaneously intruding into the preceding syllable and shortening the vowel. This has the advantage of removing the objection against violating the Onset Principle. However, Kahn's environments for ambisyllabicity, like the environments for flapping, are not the same as Myers'. If ambisyllabicity causes shortening in preceding syllables, we would again have difficulty explaining commonplace level I exceptions like motor, total, votive, not to mention failures to shorten seen in other strata across words and phrases with ambisyllabic onsets, e.g. coating, eat another potato. Once the other evidence for

---

6 Myers' constraint requires all feature matrices to be associated with a timing unit (1987: 515). His shortenings involve deletions of timing units rather than feature matrices.

7 Although one could also argue that the vowel was shortened and then re-lengthened after deletion of /h/.
Resyllabification proves to be less than persuasive, we are left with a rule which shows no effects apart from the shortening which it was devised to explain.

An objection to another part of Myers' account is found in Kager (1989:119). He notes that while Myers attempts to characterize the stress shift needed to explain forms like déference from défér through the "Sonorant Destressing" rule of Kiparsky (1979), this is not supported by the facts. Myers has modified Kiparsky's rule in two significant ways: a long vowel is treated as a "sonorant" for the purposes of the rule, and primary stress can be removed. Kiparsky's principal examples, such as infirmary, dispensary, contradict the second claim. But the first is clearly violated by forms with real sonorants, such as dependent, fraternal, as well as long-vowel forms like reprisal. The fact that no -ent /ant words with true sonorant consonants in the relevant position show this destressing is a clear indication that "Sonorant Destressing" is not responsible for the déference type. In fact, only roots with long vowels in their unstressed forms undergo this type of stress shift and shortening.

2.4 The foot as a prosodic category

Myers' account was constrained by the prosodic theory he adopted from Selkirk (1984), which focused upon the syllable as the relevant rule domain. Such a framework favors explaining segmental changes via adjustments within the syllable, as in the Resyllabification rule discussed above. Halle and Vergnaud (1987), Burzio (1993) and Prince (1990) acknowledge Myers' generalizations about the distribution of vowel alternation, but recast the focus environment for the rule in terms of another prosodic category, the foot.

The foot was introduced into current metrical theory by Liberman and Prince (1977), and in various forms has been used as a category intermediate between the syllable and word in Selkirk (1980), Hayes (1982), Halle and Vergnaud (1987), McCarthy and Prince (1990), among others. Its utility as a prosodic category has been demonstrated by, among other things, the morphological use of the foot in reduplication and infixation processes as seen in McCarthy and Prince (1986, 1990), as well as by stress-related properties summarized in Hayes (1995). McCarthy and Prince also argue persuasively for the role of the foot in defining the domain for quantitative constraints, such as the Bimoraic Constraint seen in Arabic and other languages. Myers' description of a stressed syllable followed by an unstressed syllable exactly matches the structure, for example, of the bisyllabic trochee of Hayes (1995), one of his permissible foot shapes.
If we regard the foot as a potential rule environment, it is possible to better characterize, for example, the phenomena explained above using Borowsky's Resyllabification rule in terms of the foot. The generalization Borowsky and Selkirk had attempted to capture with their syllable-based rules was extended directly from Kahn's insights, which first articulated, in the context of generative theory, the need for certain rules to refer to syllabic (and therefore prosodic) structure. However, in the cases mentioned above, the prosodic constituent which best describes the domain for these rules is, rather, the foot, a theoretical entity not available to Kahn.

Regarding these rules, /y/ can now be seen as deleted in the foot-onset, bleeding the assimilatory palatalization which may occur elsewhere in the foot. Aspiration is shown to be a feature which marks both the foot and word onsets. The rules can be restated more simply as follows:

(13) aspiration: preserved in the onset of the foot or prosodic word
    ve (hicul) <ar> vs. (vehi)<cle>
    pro(habit) vs. (prohi)biti<on>

  y-deletion: occurs after a coronal in the onset of the foot
    an(nuit)<y> vs. (annu)<cal>
    vo(lumin)<ous> vs. (volume)

Thus, processes previously explained by a questionable "Resyllabification" rule can now be represented as conditioned by position within metrical constituents.

2.5 Foot-based accounts of vowel alternation

Halle and Vergnaud (1987: 252-4) do not devote much space to vowel alternation, but, rejecting Resyllabification, regard the shortening effect as occurring in two separate environments: in closed syllables and in "the heads of binary metrical constituents", i.e. feet. They see a parallel between the two cases in that both shortenings take place inside metrical constituents, either the foot or the syllable. Burzio (1993) also recognizes that vowels appear to shorten in bisyllabic feet. However, he views this as part of a general principle (with many exceptions) that "vowels shorten in word formation" (p. 364),

---

8 It should be clear from the flapping facts noted above that flap should not be treated as parallel with the other rules.
9 Below, feet are denoted by parentheses. Feet not relevant to the discussion are not marked. Extrametrical syllables are marked with angle brackets, as above.
extending Kiparsky's (1982a, b) notion that underived forms are exempt from structure-changing cyclic rules.

Prince (1990: 367-70) provides a clearer motivation for the alternation, regarding the shortening seen in (2-3) above as the result of a tendency to optimize foot shape. As the "best" trochaic foot shape, under his definition, is composed of two light syllables, the less optimal shape "heavy-light", as seen in possible forms like *coːnic, is shortened, resulting in the better "light-light" pattern of attested conic. For monosyllabic feet, the heavy syllable of the type cone is the only possible foot, as the foot in such a quantity-sensitive system is defined as minimally composed of two units of weight, or moras (see McCarthy and Prince 1990; Hayes 1995). Prince thus improves upon Myers' account, retaining the scheme of a common environment for types (2-3) while avoiding problematic rules like Resyllabification. Utilizing the concepts of quantitative constraints and the optimization of foot-shape, he is able to present the various vowel alternations as natural and well-motivated.

Yet Prince leaves a number of questions unanswered. While the foot is undoubtedly the domain for these adjustments, it is not clear how all the attested feet would be laid down, especially in cases like (3) where a stress shift from the unsuffixed form is not explained. Additionally, the mechanism behind exceptional forms like notify and basic is not covered; presumably, these ill-formed feet are exceptionally maintained. In fact, there are further generalizations about the English data which lead us to question certain typical assumptions about the form of the lexical entries used in these derivations.

3. Revising lexical entries
3.1 Vowel length distribution in the Latinate English lexicon
The items discussed above have been at times described as [+ Latinate] in the literature, a diacritic which served to capture the observation that these words, primarily stemming from Latin historically, share common characteristics that distinguish them from other English words. The "Latinate" lexicon will be defined here as including all lexemes undergoing suffixation with the various level I Latinate suffixes (such as -or, -ar, -al, -ary, -ous, -ory, -age, -ance, -ence, -esce, -ic, -ate, -ion, -ment, -ive, -ant, -ent, -ure), regardless of their historical origin.\(^{10}\)

\(^{10}\) While Myers treats the irregular preterite tense marker -\textit{d/} and the Anglo-Saxon suffix -\textit{th/} similarly, I regard this similarity as superficial; in any case, none of the lexemes taking these suffixes have any interaction with the Latinate suffixes.
The class of lexical items which appear with Latinate suffixes like -ic, -al, -ity, -ent noted above can be described as "Latinate stems". The term "stem" is used here to represent both the "bound roots" and "stems" of Inkelas (1989). Stems combine both with prefixes and suffixes, and do not necessarily have a clear semantic locus, e.g., -ceive, which is "bound" in that it never appears without various prefixes, as in receive, deceive, conceive, but can be formally regarded as a single lexeme due to the distribution of clearly distinguishable prefixes in the lexicon as well as common features of allomorphy throughout the -ceive paradigms, for example the extended form in -cept, concept, deception, perceptive.

Herein, the Latinate stems will be regarded as a grammatically defined class of stems sharing not only word-formation processes, but a distinct set of rules and constraints which do not apply to other stems of English. Additionally, the Latinate stems share further distributional characteristics: for example, the vowel /U/, not uncommon in English (e.g., book, hoof, pull) is virtually absent from the Latinate lexicon, while /U/ may only appear initially in Latinate stems (e.g. re-fer, de-flect), never in other positions, as in raft, wafer, scoff\textsuperscript{11}.

While Myers noted that long vowels in English were mostly distributed into open syllables, motivating his use of a syllable template, he failed to notice another, widespread generalization: among the Latinate stems, monosyllabic light stems (closed by a single consonant) overwhelmingly show long vowels, while monosyllabic heavy (closed by two consonants) and multisyllabic stems overwhelmingly show short vowels, in their unsuffixed forms\textsuperscript{12}. The same facts hold for both "bound" and "free" stems:

\begin{tabular}{llll}
(14) & Monosyllabic Open & Monosyllabic Closed & Multisyllabic \\

tone & /to:<n>/ & act & /ak<>/ \\
globe & /glo:<b>/ & se-lect & /-lek<>/ \\
com-bine & /bi:<n>/ & de-fend & /-fen<d>/ \\
de-ceive & /-ce:<v>/ & inter-rupt & /-rup<>/ \\
con-done & /-do:<n>/ & sub-sist & /-sis<>/ \\
(etc.) & (etc.) & (etc.) & (etc.)
\end{tabular}

\textsuperscript{11} For further justification of this approach, the reader is referred to my forthcoming dissertation.

\textsuperscript{12} Here, important previous assumptions about the grammar are maintained: phonological long vowels are reflected in surface diphthongs; word-final consonants are regarded as extrametrical, following Hayes (1986).
Only a very small number of commonly used Latinate stems do not fall under this generalization:

(15)  add  ab-err  ob-sess  in-still  cipher
      ad-mit  re-fer  pro-gress  in-stall  libel
      dis-cuss  re-cur  con-fess  re-bel  danger
      al-lege  pro-pel  im-press  ex-cel

In fact, from a list of over 7,000 common Latinate words, 88% of monosyllabic bare stems showed a long stem vowel, while 94% of multisyllabic bare stems showed only short stem vowels\textsuperscript{13}.

This is far from what we should expect by chance. It has further ramifications, because of an assumption about the lexicon that goes back to Chomsky and Halle (1968) and its structuralist ancestors: unsuffixed forms are assumed to be basic, that is, they are assumed to reflect the underlying lexical entry for the stem in question. Thus, it appears that most monosyllabic Latinate stems show underlying long vowels, while most multisyllabic stems lack them. If we regard lexical short vowels as the unmarked case, this part of the grammar of English unexpectedly shows tremendous markedness, since, among monosyllabic lexemes, putatively marked entries far outnumber unmarked ones. Yet, this markedness disappears in the case of the multisyllabic stems.

Furthermore, comparing these Latinate lexemes to the set of English words that do not undergo Latinate word-formation (e.g., hit, bat, paint, lead, foot, set, walk, etc.), in which there does not seem to be any clear preference for short or long vowels in the unsuffixed (and presumably underlying) consonant-final stems\textsuperscript{14}, suggests that something unusual is happening with regard to the forms of the Latinate set: the appearance of long vowels in stems ending in a single consonant (which can be regarded as syllables "open" via extrametricality) seems to be dependent upon the number of syllables found in the stem itself.

\textsuperscript{13} This list was compiled using a Unix-based dictionary as well as conventional backwards dictionaries. A more exhaustive, fully computational study of a 50,000 lemma list is being undertaken for the purpose of further precision. Preliminary results suggest similar proportions will be found. The words in (15) represent an exhaustive sample of exceptional unsuffixed bases from this list.

\textsuperscript{14} There is, however, a similar distributional effect that will be remarked on below: while there is no pattern of vowel length distribution in consonant-final monosyllabic "native" roots, monosyllabic vowel-final roots uniformly show long vowels. That is, we have go /go/, be, /be/, die, /di:/, etc., but no */go, *be, *di/ or the like.
3.2 Minimality constraints and moraic weight

A solution to this question of distribution can be found in the concept of a minimality constraint, mentioned above in the account of Prince (1990). Minimality has been used to explain quantitative adjustments based on prosodic domains. McCarthy and Prince (1986) propose minimal word and foot sizes to account for a number of morphological processes cross-linguistically. A minimally constrained prosodic category, the bimoraic foot, forms the basis for the moraic trochee of Hayes (1995), which underlies quantity-sensitive foot-formation. That is, feet in languages of this type must consist of two moras, or else be "degenerate". Similarly, in certain of these languages, empirical evidence suggests that prosodic words must consist of at least one well-formed foot.

In many languages, items failing to meet the criteria for well-formedness on their own are moraically altered to satisfy the minimality requirement. For example, as discussed by McCarthy and Prince (1990b: 21), in Modern Standard Arabic, vowels in monomoraic loan words are lengthened to achieve the bimoraic limit, e.g., baar 'bar', jaaaz 'jazz', gaaz 'gas'. A Saudi Bedouin dialect geminates final consonants in the same situation, e.g. bass 'bus', natt 'nut', rigg 'rig'. In a different type of case, the reduplicative prefix in Mokilese is obligatorily a bimoraic syllable, regardless of the moraicism of the initial root syllable. Thus, light roots show a lengthened vowel in the reduplicative prefix; wad-wadék 'read', pok-poki 'beat', and caa-caaak 'bend' might appear to simply repeat the form of their initial syllables, but paà-pa 'weave' and wifi-wia 'do' show the lengthening required by the minimality constraint (McCarthy and Prince 1986: 20-1).

While the bimoraic trochees proposed for English foot-formation (e.g. Hayes 1989; Hayes 1995) are usually assumed to be built around heavy syllables found underlyingly in the lexical entries of words like tone, minimality and the concomitant lengthening which may enforce it allow us to propose an alternate hypothesis: Latinate stems of the "open" type in (14) above, i.e. those underlying tone, globe, conceive, condone, etc., are **underlyingly short**, containing only one mora, i.e. /ton/, /glob/, /cev/, /don/. To fulfill

---

15 Degenerate feet are subject to a number of restrictions, if they are allowed at all (see Hayes 1993: 75-9).
16 Other accounts, like Halle and Vergnaud (1987) or Kager (1989), assume that such "underlyingly long" vowels have "intrinsic" prominence, rather than forming the basis for feet. The assumptions and the effects are, with respect to the issues discussed here, identical to those in Hayes' approach.
17 While Prince (1990: 370) cautiously suggests that some vowels in similar cases might be "unspecified" for length, and Kager (1989: 119) suggests positing underlying short vowels in other cases (he regards these as plausible only in pairs that have lost their derivational connection, e.g., confide : confident), to my knowledge no other analysis has proposed this idea of lengthening rather than shortening to explain this alternation.
a minimality constraint, their stem vowels are lengthened to create a bimoraic foot. This can be illustrated using the proposed stem /ton/\(^{18}\):

<table>
<thead>
<tr>
<th>Lexical Entry</th>
<th>Minimality</th>
<th>Surface bare form</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ton/</td>
<td>(to)&lt;n&gt;</td>
<td>tone (to:&lt;n&gt;)</td>
</tr>
</tbody>
</table>

This stem is the only phonemic input to the formation of the foot (ton). Remembering that word-final consonants are extrametrical, we thus have a foot (to:<n>), with only one mora, that of the vowel. The minimality constraint forces a lengthening to (to:<n>), which may be conceived of as the linking of another mora to the foot. The vowel is the only viable linking point, as neither the onset nor the extrametrical consonant can link to the mora (Hayes 1989). The short surface vowel seen in tonic is then not the result of a shortening, but instead reflects the underlying short quality of the stem vowel. Because -ic is not extrametrical (see Hayes 1982), it may join with the stem /ton/ in foot formation, i.e. (toni)<k>, the final consonant of the word remaining, as usual, extrametrical:

<table>
<thead>
<tr>
<th>Lexical Entry</th>
<th>Minimality</th>
<th>Surface bare form</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ton/ + /ic/</td>
<td>(toni)&lt;k&gt;</td>
<td>tonic (toni)&lt;k&gt;</td>
</tr>
</tbody>
</table>

3.3 A reanalysis of exceptional stems

Understanding the stem of the common tone type as monomoraic now admits a further reanalysis of the underlying lexical forms for Latinate stems. As will be illustrated below, all types are now able to be represented without diacritic exception marking; only the exceptional appearance of a long segment, marked in the lexicon by an extra mora, distinguishes the different monosyllabic stem classes:

\(^{18}\) Here, feet are delimited by parentheses, while extrametrical material is enclosed in angle brackets.
(18) | **Lexical Entry** | **Surface bare form** | **Affixed form** |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>/ton/</td>
<td>tone (to:) &lt;n&gt;</td>
<td>tonic (toni) &lt;k&gt;</td>
</tr>
<tr>
<td>/not/</td>
<td>note (no:) &lt;t&gt;</td>
<td>notify (nò:ti) &lt;fy&gt; ¹⁹</td>
</tr>
<tr>
<td>/add/</td>
<td>add (ad) &lt;d&gt;</td>
<td>additive (áddi) &lt;tiv&gt;</td>
</tr>
</tbody>
</table>

This proposal eliminates the unexpected markedness and complexity inherent in a lexicon that appeared to show, nearly universally, underlying long vowels in monosyllabic light stems. Instead, the vast majority of stems are seen to be monomoraic, the unmarked type.

Additionally, there is now a principled way of marking exceptional stems, like the /nòt/ of note, notify: which apparently fails to undergo the expected "shortening" proposed in prior analyses. Instead of resorting to unstructured exception marking, these infrequent stems can be understood as underlyingly long (unlike the majority of "open" stems, such as /ton/). Words like notify fail to undergo "shortening" because there is no longer a rule of shortening, but rather one of lengthening in monomoraic words, to achieve a minimal foot. As /nòt/ and other stems like it are already bimoraic, they are simply unaffected, and surface with long vowels in all forms. Following this account, stems showing underlyingly the more marked long vowels are seen to present a very limited distribution in the lexicon, rather than the unexpected ubiquity they apparently display under the standard proposal.

The few monosyllabic "open" stems with short surface vowels, such as those of add, sub-mit, re-pel, re-gress can also be accounted for similarly. If they are re-analyzed as underlyingly geminate²⁰ and so also bimoraic, i.e., /add/, /mitt/, /pell/, /gress/, the minimality constraint will be satisfied by the stem itself, and no lengthening need take place. These stems with geminate consonants represent another marked, and

---

¹⁹ Neither this form nor (ad:it)<iv> below it seeks to take a side in the debate over whether such words should be parsed as such or as (ad:)it<iv>, with non-footable segments being "skipped over" (see Hayes 1993; Kager 1989). I do suggest that Prince's (1990) rule that shortens long segments to avoid such problematic sequences is not operative.

²⁰ Interestingly, all these proposed geminated stems end in coronals. Underlying geminates have also been proposed by Burzio (1993: 365) to explain the stress seen in words like vanilla.
distributionally limited, type; geminates can be expected to reduce at the surface level in English, as when a compound like roommate, /ruːm/ + /maːt/, shows up as [ruwmeyt] rather than as *[ruwmmeyt] with a surface geminate.

This leaves the vast majority of vowels in the Latinate lexicon, across all three of the stem-types noted in (14), to the unmarked case, underlying short vowels. Deviations from this type are marked with an exceptional extra mora, rather than a diacritical exception marker. The concept of foot and word minima also allows for an explanation of the distribution of vowel length in non-Latinate words, remarked upon in a footnote above. The fact that no words of the form /*be/, /*go/, /*di/ occur suggests that the same bimoraic constraint holds for the "native" grammar as well. The difference is that final consonants, extrametrical in the Latinate grammar, appear to hold a mora in the native lexicon, i.e. (hit), (bet), rather than *(hi:)<t>, *(be:)<t>. The contrast between the two systems may hinge upon nothing more than this difference in final consonant extrametricality.\(^{21}\)

3.4 Stress shift and foot-formation

The foot-based analysis and reinterpretation of the underlying forms of Latinate stems proposed above also allows for the approach of Prince (1990) to be improved upon. The requirement that feet should be bimoraic is now reflected by a lengthening process similar to rules seen in, for example, McCarthy and Prince (1986, 1990). Furthermore, Prince's observation that bisyllabic bimoraic feet are the optimal trochees is now mirrored in the complex of the lexicon, rather than by a shortening rule which would not always take effect, as before. This allows for the infrequent cases, such as danger, notify, which unexpectedly show long vowels. But more significantly, the new characterization of the "open" roots such as tone as underlyingly monomoraic helps to clarify cases of foot-formation which were inexplicable under the former analysis, i.e., the words of set (3), as noted above.

As long as the stems underlying provide / provident and other set (3) words are assumed to contain long vowels, there is no way to produce the foot structure (provi)<dent> needed for Prince's shortening rule\(^{22}\). Instead, *pro(vi)d<ent>, parallel to re(spond)<ent>, would be expected, as a syllable with a long vowel and thus "heavy" or

\(^{21}\) This also implies that, in the face of examples like width, dreamt, there is a maximality constraint in the "native" grammar, resulting in true shortening, that is not present in "Latinate" English. This shortening seems confined to certain unproductive derivational forms.

\(^{22}\) Myers' explanation of this type, involving a destressing rule, was discussed above, along with Kager's (1989) argument rejecting it.
bimoraic, to comprise a foot in itself and receive main stress. However, if the stem /-vid/ is understood as underlyingly monomoraic, the difference between these words can be easily accounted for:

(19) \( (\mu) (\mu\mu) <(\mu\mu)> \)  \( (\mu \mu) <(\mu\mu)> \)  
\begin{tabular}{ll}
re - spond - ent \\
pro - vid - ent
\end{tabular}

Following Hayes (1986), the suffix /-ent/ is extrametrical; the root-final /d/ in both cases will syllabify as the onset of this extrametrical onsetless suffix. Thus, foot-formation begins with the stem syllable in both cases. The stem /-spon<\(d>)/ is intrinsically bimoraic, and can form its own foot, taking on the main stress. However, /-vi<\(d>)/ is monomoraic, unable to comprise a foot on its own. If the prefix /<pro>-/ is similarly analyzed as monomoraic, both potentially degenerate feet can together form the optimal bisyllabic bimoraic foot (prôvi). The stressed vowel, as predicted, is short. Conversely, the monomoraic prefix /re/ is excluded from the already bimoraic foot (spond) in respondent.

Here, the two possible treatments of the prefix in respondent illustrate the advantage of proposing a monomoraic underlying form for the traditional long-vowel prefixes, e.g. /pre/, /rel/, /pro/. One variant shows a lengthened, stressed prefix vowel, that is, the prefix forms a foot of its own; to be well-formed, it must also become bimoraic, thus (rè:)(spônd)<ent>. The other variant instead shows an unstressed prefix, with reduced vowel. Here, the prefix is excluded from any foot, and its vowel is instead reduced, like the vowel of the extrametrical suffix /-ent/: [rə(spônd)<ənt>]. This analysis accounts not only for the "shifted" stress in words with monomoraic prefixed roots with heavy suffixes like /-ent/, as in prôvident, but also explains, again through using the bimoraic foot, the various forms that the traditional long-vowel Latinate prefixes may yield in these cases. The usual interpretation of both the "open" stems and the "open" prefixes as underlyingly containing long vowels is unable to account for these facts.

3.5 Derived and non-derived forms

This re-evaluation of stems like /hon/ as underlyingly short presents further theoretical ramifications. The framework of Chomsky and Halle (1968), like those that both preceded and followed it, assumed that unsuffixed forms like tone, pose, were identical to the underlying representations of their stems, / tôn/, /pôs/, thus the general interpretation of forms like tonic, posit as reflecting shortening. While generative theory is known for introducing far greater abstractness into underlying representations (Anderson 1985:}
the treatment of many words still reflects the (intuitively appealing) practice of the earlier structuralists, who regarded only surface forms as having a linguistic reality (see Anderson 1985: 306-9). This assumption was taken over into Lexical Phonology, where it was strengthened and codified into the Alternation Condition of Kiparsky (1973) and the Strict Cycle Condition of Mascaró (1976). These principles stated that cyclic (for some, lexical) rules could apply only in "derived environments", that is, only to words undergoing affixation processes.

Such an assumption would prevent lexical rules from applying to words like tone and pose, and might thus appear in conflict with the proposed analysis. There are a number of responses to this objection, however. Later work, such as Kiparsky (1982b) exempted metrical structure rules, such as syllabification and stress assignment, from this restriction. Thus, a foot-based rule like our constraint-based lengthening, which could be classified as "structure building"\textsuperscript{23}, might also be expected to apply to underived forms. In fact, most examples of moraic lengthening due to minimality requirements, as described in the literature noted above, apply in underived environments. That notwithstanding, there are further reasons to question the elevation of derived versus non-derived status to such a fundamental position in both the grammar and the theory, as suggested by the Alternation Condition and its successors. There are many languages, for example Latin, in which virtually all lexical words show inflectional morphology. In such languages, the supposedly fundamental grammatical distinction between derived and non-derived forms suggested by Lexical Phonological theory is, for unexplained reasons, virtually absent.

Furthermore, there exist in Latin isolated cases of apparently "non-derived" words, which, for historical reasons, have lost their surface case marking. For example, puer (nominative) puer-is (genitive) 'boy', pater (n.), pater-is (g.) 'father' appear to show an unsuffixed stem as the nominative form, as opposed to regular dux (n.), duc-is (g.) 'leader', morphologically /dʊk-s/, /dʊk-is/, /s/ being the regular marker of the nominative singular. To take the example for 'father', there is no valid reason not to posit an underlying /patr/ which undergoes an /e/-insertion in the "basic" nominative, rather than an underlying /patr/ which apocopates the vowel in the derived forms. In fact, there are counterexamples to such a proposed rule of apocope, e.g., soror, sororis 'sister', showing the long vowel in both environments. The rule producing the paradigm for 'father' must be lexical, as neither the proposed insertions nor deletions are regular.

\textsuperscript{23} Kiparsky's exemption is based on the empirical observation that rules, such as stress assignment, which add structure to, rather than alter, the underlying lexical string clearly may apply to all lexical words.
Moreover, the same situation arises in English, in a few pairs like sign / signature, damn / damnation. For these, Kiparsky (1985) suggests a relaxing of the constraint on the last lexical level, a theoretical tactic described by Kaisse and Shaw (1985: 23) as "disappointing". In fact, the original reason for distinguishing between derived and non-derived environments must be recalled: the failure of words like nightingale to undergo Chomsky and Halle's (1968) rule of Trisyllabic Shortening (cf. Kiparsky 1982: 35-42). When unsuffixed forms are no longer assumed to be identical to underlying forms, and vowel alternation is seen as a result of lengthening, rather than shortening, it becomes clear that a rule of Trisyllabic Shortening is no longer necessary. Rather, Trisyllabic Shortening is best seen as a generalization, applying only to a subset of words, describing the effects of a prosodically-based constraint which was simply not available theoretically to Chomsky and Halle (1968). Accepting this, the problem of explaining why Trisyllabic Shortening and similar generalizations fail to apply to certain forms vanishes as well.

4. Conclusion

By representing what have been previously understood as monosyllabic long vowel roots instead with underlyingly short-vowel bases, both the surface long vowels (lengthened due to foot minimality constraints) and short vowels (maintained due to their membership in already bimoraic feet) found in most Latinate words can be better accounted for, ridding the system of problematic shortening rules, and subsequently, of further theoretical baggage needed to support those rules, such as the derived/underived dichotomy and various destressing rules. The remaining cases of unexpected long and short vowels in Latinate stems may then be understood as marked for length in the lexicon, rather than marked as exceptional to certain rules. An otherwise inexplicable skew in the surface frequency of long vowels, correlated with stem length, becomes understandable in light of ideas concerning both the minimal and optimal sizes of prosodic constituents.

My approach assumes that it is better to represent exceptional forms by adding structural information to their lexical entries, in this case, an extra mora, rather than by marking them as exceptional to undergoing certain rules. Such "unstructured" exception marking is seen as potentially unconstrained, while exception marking utilizing structural mechanisms limits the addition of structure onto lexical items to demonstrably necessary constituents. Also, by recording idiosyncracies directly on the lexical entry, all lexical items may be made to undergo the same set of rules.
This preliminary study only touches upon a few types of English word-formation processes, albeit those at the heart of the problem of vowel alternation and suffixation. Of course, there are further cases which require more elaboration and discussion before they can be regarded as similarly tractable. Additionally, this work does not take into account the potential contributions of Optimality Theory (McCarthy and Prince 1993; Prince and Smolensky 1993), which, with its focus upon constraint-driven phonology, offers a framework which may further facilitate a more complete explanation of English word-formation along the lines suggested here. These problems call for further study, and for this the reader is referred to my forthcoming doctoral dissertation, which will treat these issues in greater depth and breadth.

5. References


