Degenerate feet in phrasal phonology: evidence from Latin and Ancient Greek*

Abstract

Degenerate feet, even when forbidden in isolated words, can arise within phrases due to resyllabification. In particular, when a stressed monosyllable of the shape $C_0VC$ (where $V$ is short) undergoes resyllabification in Latin and Ancient Greek, it yields a degenerate foot. While degenerate feet were tolerated in prose, they were avoided in hexameter verse. Even though a degenerate foot is a kind of light syllable, a light metrical position could not contain a foot. Verse evidence is used as a window onto the general prosodic structure of each language, revealing that speakers productively recognized degenerate feet and distinguished them from other prosodic categories.

Degenerate (that is, monomoraic) feet, while permitted in some languages, tend to be highly restricted in their distribution (Hayes 1995). In Ojibwe, for instance, a degenerate foot is possible only at the end of a prosodic word (PWd), as in (1) (Newell and Piggott 2014), suggesting that PARSE, which requires syllables to be footed, dominates FTBIN, which requires feet to be binary. (Throughout, feet are parenthesized.) Medial degenerate feet, as in ungrammatical (c), are gratuitous because the feet could be combined, as in (b), satisfying PARSE without compromising FTBIN.

(1)  a. $(\text{wi:})(\text{kwa:})(\text{bo:})(\text{zo})$ “he is carried along by the current”
     b. $(\text{gin\textquoteleft wa:})(\text{bi\textquoteleft ki})(\text{zi})$ “it is a long metal object”
     c. *$(\text{gin\textquoteleft wa:})(\text{bi})(\text{ki})(\text{zi})$

As I argue here, some languages that forbid degenerate feet in isolated PWds permit them in phrasal phonology, where multiple PWds interact. Classical Latin and Homeric Greek furnish two cases, along with the metrical diagnostics to confirm that the syllables in question are in fact degenerate feet, that is, both light (monomoraic) and independently footed. Specifically, degenerate feet arise in these languages due to resyllabification. An isolated PWd can be $C_0VC$ (throughout, $V$ is short), as with Latin dat “gives” in (2a). As the subscripts indicate, the PWd is bimoraic, satisfying FTBIN; indeed, isolated PWds are required to be at least bimoraic in Latin (Mester 1994) and Ancient Greek (Blumenfeld 2011). When prevocalic, however, as in (2b), dat loses its coda, which is now parsed as the onset of the following PWd. Latin, like Greek, tolerates the resulting degeneracy (subminimality). The metrical evidence reviewed below confirms that in configurations like (b), the consonant is resyllabified (without ambisyllabicity), the vowel remains short, and the first word retains its stress.

(2)  a. $(\text{da}_{\mu}\text{t}_{\mu})$ “gives”

*Acknowledgments.*
Not all languages with bimoraic minima and resyllabification tolerate resyllabification-induced degeneracy like Latin and Greek do. Tamil, for instance, geminates the final consonant (if geminable) of a C_0VC word when it undergoes resyllabification, as in (3a–b) (cf. Italian; Chierchia 1982). Crucially, in Tamil, such gemination is absent when FTBIN is not at stake, as with aval “her” in (c), which remains bimoraic after losing its coda (Ryan 2019).

(3)  
   a. (‘n aµ́lµ)  “good”  
   b. (‘n aµ́lµ)(‘l uµμfµ) “good city”  
   c. (‘aµ v aµ)(‘l uµμfµ) “her city”  

Another strategy to avoid resyllabification-induced degeneracy is to suppress resyllabification if it threatens binarity (e.g. by ranking FTBIN over ONSET). Ryan (2019) documents this situation in Māhārāstrī Prakrit, which records word-final nasals differently depending on whether or not they undergo resyllabification. The final nasal of a C_0VC word, where degeneracy is threatened, is significantly less likely to resyllabify than the final nasal of a longer word, where degeneracy is not at stake. Of course, a language might also forbid cross-PWd resyllabification across the board, as with Dutch (Booij 1996).

A final logically possible strategy to avoid resyllabification-induced degeneracy is to suppress stress. In such a language, a C_0VC word, say, tat, would be stressed when preconsonantal but unstressed when prevocalic, as schematized in (4). I am not aware of any case exactly like this one, though Hayes (1995: 112–3) considers possible cases of subminimal words being unable to surface with stress in any context.

(4)  
   a. (‘t aµ́ tµ́)(‘b aµ g aµ) (constructed)  
   b. t aµ́ (‘t aµ g aµ) (constructed)  

Latin and Greek, as mentioned, do not employ any of these strategies to circumvent a degenerate foot when one arises due to resyllabification. Rather, they let the degenerate foot stand. A benefit of examining such languages is that their extensive and exacting metrical traditions furnish independent diagnostics of prosodic properties such as weight and stress. (Whether Ancient Greek had stress in addition to pitch accent is irrelevant here; either way, I assume, along with most analysts, that it had feet, as discussed in §5.)

For both languages, I focus on the (dactylic) hexameter, as schematized in (5). The following basic description of the meter will suffice for all present purposes. Each of six metra, as enumerated, is divided into two metrical positions. The first, _, must be filled by a heavy syllable and the second, ≉∥, by either a heavy or pair of lights. If the latter is filled by a pair of lights, each light occupies a SUBPOSITION, in Prince’s (1989) terminology. Note that “subposition” is equivalent to saying “light-requiring position” for this meter, since light syllables are not otherwise licensed outside of line end. The cadence comprises the final two metra. In Latin, the cadence, unlike the rest of the line, is strictly regulated for stress, being nearly always (for most authors) stressed in strong ( _) positions and unstressed in weak positions (≈∥ and ≈).
As I will show, stressed C_0VC retains its stress even while undergoing resyllabification, resulting in a degenerate foot. While such feet were freely tolerated in prose (and thus everyday speech), they were strongly avoided in most authors’ hexameters. I attribute this avoidance to a maximum size constraint on the subposition, such that it cannot contain a prosodic constituent higher than the syllable. A degenerate foot, in effect, is too large for a subposition, but too small (being light) for a heavy-requiring position. It therefore finds no comfortable place to sit in the meter, even while being free to occur outside of verse.

To be clear, this analysis does not suppose that degenerate feet are somehow confined to poetic language or a property of the meter itself. Rather, degenerate feet are a property of the general language. Poets, in turn, can be sensitive to them, just as they are sensitive to moraic structure and other properties of the general language. In this way, verse sheds light on general prosody. Poets’ special treatment of degenerate feet, as distinct from similar structures such as stressed light syllables more generally, supports the psychological reality of the construct. I conclude that degenerate feet are more widely distributed in the world’s languages than previously assumed.

I begin with the empirical facts in Latin in §1 before turning to the formal analysis of degenerate feet as induced by resyllabification in §2. The next two sections §§3–4 address the verse evidence for degeneracy in Latin, including (in §4) a cline of stressability among function words. I turn to Greek in §5, demonstrating that essentially the same behavior is found in the Homeric hexameter. Finally, §6 considers alternatives to the degenerate foot, with §7 concluding.

1 Resyllabification-induced degeneracy in Latin

Latin is a classical example of a moraic trochee language (Mester 1994). Feet are initially prominent and may be two syllables or one heavy (bimoraic) syllable. Degenerate feet are forbidden in isolated words, as suggested, among other evidence, by prosodic minimality (6a), stress placement (6b), and, especially in Early Latin, trochaic shortening (6c). Thus, in Optimality Theory, foot binarity (FTBIN) is ranked highly in Latin (Prince and Smolensky 1993/2004).¹

(6) a. /da/ → (‘daː), *(‘da) “give”
  b. *(ani)ma, *(‘a(ni)ma “soul”
  c. /amo:/ → (‘amo), *(‘a)mo: “I love”

That said, degenerate feet, I argue, arise in cross-word contexts. In Latin, as in many of its descendants, a word-final consonant resyllabifies with a following vowel-initial word (Allen 1978, Harris 1983, Peperkamp 1997, Ryan 2019, Passino et al. 2022).² If a stressed monosyllabic of the shape C_0VC is prevocalic, resyllabification yields a degenerate foot, as in (7a). As argued presently, such phrases are not normally parsed as (7b) (suppressing resyllabification), (7c) (lengthening the vowel), or (7d) (suppressing stress), among other possibilities (the possibility of

¹Encitics, including que “and,” induce stress on a stem-final light in at least some contexts, as in sceleraque (probably also with stress on sce). If foot-based NONFINALITY were undominated, such forms would yield a degenerate foot. However, as (6) has already suggested, NONFINALITY is not undominated. Mester (1994) suggests one analysis of enclitics that obviates degeneracy, though the empirical situation is insecure (cf. Newcomer 1908, Allen 1978). At any rate, in agreement with Mester (1994), there is no strong case for degenerate feet in the context of enclitics. Even if there were, the issue is orthogonal to the treatment of phrases that follows.

²Pronunciation, not orthography, matters. For example, h does not block resyllabification in Latin or Greek.
ambisyllabicity, for one, is rejected in §

(7)  *dat Aulus  “Aulus gives”
   a. (*‘da’) (‘t au)lus
   b. *(‘dat) (‘au)lus
   c. *(‘da:) (‘t au)lus
   d. *da(‘t au)lus

First, resyllabification applies as normal to C₀VC words. Among other evidence, this is clear from verse, where such words scan as light prevocally. In the lines in (8), for example, *vir “man,” *fer “carries,” and *dat “gives” must scan as light to fit the hexameter. If prevocalic C₀VC resisted resyllabification, as in (7b) above, one would expect it to be treated as heavy in verse.

(8)  a. *Vāde age, et ingentem factīs *fer ad aethera Trōiam (Virgil, Aeneid 3.462)
   b. *Hic *vir, hic est, tibi quem prōmittī saepius audīs (Virgil, Aeneid 6.791)
   c. *prīncipium dat et hinc mōtūs per membra rigantur (Lucretius, De rerum natura 2.262)

Second, C₀VC does not undergo lengthening of the vowel or consonant when prevocalic, which would yield a heavy syllable. This again is clear from scansion, as in the verse lines just provided in (8), as well as from manuscripts in which vowel length is indicated.

Finally, lexical C₀VC does not lose its stress due to resyllabification, as in (7d) above. For starters, to my knowledge, no language with phrasal resyllabification works that way. In Italian, for instance, *snob horrendo [ˈsnɔ.b ˈhɔr.rento] “horrible snob” exhibits resyllabification, but the monosyllable retains its stress, as reinforced by [ɔ], which diagnoses stress (Peperkamp 1997). In Latin, too, resyllabification does not induce prosodic fusion. This is clear, once again, from metrics. Ovid, for instance, virtually requires strong positions to be stressed in metrical cadences (Sturtevant 1923). A typical cadence such as *mīlisīsū [ˈmiːli.ˈsiːs.oː] “from use of the soldier” would be impossible if resyllabification induced destressing of either word.

Metrical evidence further supports the conclusion that C₀VC monosyllables in particular do not lose their stress when undergoing resyllabification. Consider once again Ovid’s Metamorphoses (8 CE), in hexameter. Recall that stressed light syllables are rarely permitted in the cadence, though they are frequently found elsewhere in the line (Sturtevant 1923, Ryan 2017). If, on the one hand, lexical C₀VC words retain their stress under resyllabification, one would expect them to be excluded from cadences. If, on the other hand, they lose their stress under resyllabification, they should be free, like unstressed function words, to occupy cadences when prevocalic.

In Metamorphoses, lexical C₀VC#V never occupies the cadence, whereas function C₀VC#V often does. As representative function words, I take all C₀VC prepositions from the text, namely, *ad “to,” *in “in,” *ob “against,” *per “through,” and *sub “under.” As representative lexical words, I take all C₀VC nouns and verbs from the text, namely, *dat “gives,” *det “gives (subjunctive),” *fac “make,” *fer “carry,” *flet “weep,” *it “goes,” *nat “swims,” *vir/lur “man,” *scit “know,” *stat “stand,”

---

3To be sure, the threat of sub minimality can trigger prosodic fusion in some languages under some conditions. For example, in Kabardian (Gordon and Applebaum 2010), a CV word such as /ʃə/ “new” is subminimal. If possible, it will fuse with a host, as in [wənʃə] “new house.” But this fusion, which is highly restricted morphosyntactically, is not induced by resyllabification.

4Ab “from” is set aside because it is usually ə before a consonant.
and *stet* “stand (subjunctive).”

Corpus counts are provided in (9). While C₀VC#V prepositions are roughly equally frequent inside and outside of the cadence, lexical C₀VC#V occurs only outside of the cadence. This significant difference between lexical and functional words (Fisher’s exact test $p < .0001$) suggests that stressed C₀VC retains its stress even when prevocalic.

(9) **Prevocalic C₀VC words in Ovid’s *Metamorphoses* in the pre-cadence vs. cadence**

<table>
<thead>
<tr>
<th></th>
<th>pre-cadence</th>
<th>cadence</th>
</tr>
</thead>
<tbody>
<tr>
<td>preposition</td>
<td>360</td>
<td>527</td>
</tr>
<tr>
<td>lexical</td>
<td>36</td>
<td>0</td>
</tr>
</tbody>
</table>

A second and independent argument that lexical C₀VC retains its stress under resyllabification concerns Virgil’s hexameter. Virgil avoids lexical C₀VC before a vowel altogether, regardless of position in the line. In Virgil’s three hexameter works, the *Eclogues, Georgics, and Aeneid* (collectively c. 37–19 BCE), lexical C₀VC (using the same lists as above) occurs prevocally six times. Preconsonantally, by contrast, it occurs 119 times. In other words, non-line-final lexical C₀VC precedes a vowel 5% of the time. This rate differs significantly from that of prepositions, which precede a vowel 41% of the time (Fisher’s exact test odds ratio = 14, $p < .0001$). Corpus frequencies are provided in (10). This difference in Virgil’s treatment of lexical and functional C₀VC arguably reflects stress: Virgil strongly disfavors stressed monosyllables in light-requiring positions, even when they are rendered light by resyllabification. A formalization of this avoidance is presented in §3.

(10) **C₀VC words in Virgil’s *Eclogues, Georgics, and Aeneid* before a vowel vs. consonant**

<table>
<thead>
<tr>
<th></th>
<th>prevocal</th>
<th>preconsonantal</th>
</tr>
</thead>
<tbody>
<tr>
<td>preposition</td>
<td>764</td>
<td>1,299</td>
</tr>
<tr>
<td>lexical</td>
<td>6</td>
<td>119</td>
</tr>
</tbody>
</table>

In conclusion, the final consonant of a C₀VC word resyllabifies when prevocalic, regardless of whether the word is lexical or functional. If the word would otherwise be stressed, it retains its stress under resyllabification. In a foot-based formalism, these conditions necessitate a degenerate foot.

### 2 Resyllabification-induced degeneracy: analysis

As a simple representative example, consider *it as* “the coin (as) goes (it).” The winner [‘(i)('t as)] in (11) exhibits a degenerate foot induced by resyllabification. (To save space, when a PWd dominates nothing but a foot, they are notated on the same line as “f, ω.”) Resyllabification is driven by **ONSET**, which penalizes a syllable with an empty onset. Because Latin permits isolated words to

---

5This list excludes any C₀VC word with a long vowel or any variant with a long vowel. For example, *os* was excluded, being orthographically ambiguous between *os* “bone” and *ös* “mouth.” Forms of the copula *esse* were not counted as lexical verbs. Words ending with *m*, such as *cum* “with” and *rem* “thing,” were excluded because final /Vm/ is subject to elision (especially in longer forms), suggesting that the rime might have at least optionally been realized as [V:]. Finally (as suggested by Brent Vine, p.c.), any word possibly ending underlingly with a cluster or geminate was excluded (e.g. *as* “coin,” *cor* “heart,” and *mel* “honey”).

6I previously made this point in [redacted] based on a smaller corpus, namely, the first six books of the *Aeneid.*
begin with vowels, ONSET must be dominated by faithfulness constraints precluding prothesis and aphaeresis (not shown).\footnote{Relevant constraints include DEP, MAX$_{\sigma_1}$, and ANCHOR.} ONSET must in turn dominate any constraint penalizing resyllabification (Ito and Mester 2009). The tableau also includes “weight by position” (WbyP: coda consonants are moraic), which dominates *µ/C (consonants must not be moraic), a ranking necessary to account for the weightiness of codas in Latin.

ALIGN in (11) is short for ALIGN($\text{lex}$, R, $\omega, R$) “the right edge of a lexical word coincides with the right edge of a PWd.” A penalty is incurred by each misaligned segment. A separate constraint ensures that lexical words are parsed into PWds: MAX($\text{lex}$, $\omega$) “a lexical word corresponds with a PWd” (Ito and Mester 2019). This constraint checks the existence of a coindexed PWd, not its alignment. While a more traditional formulation of MATCH (Selkirk 2011) might wrap these two functions (alignment and existence) into a single MATCH constraint, I follow Ito and Mester (2019) in separating them, as they must be ranked differently: Alignment is violable; existence is not. The indexation to lexical words reflects that, as Selkirk (2011) puts it, “lexical category words are standardly parsed as prosodic words ($\omega$), while functional category words like determiners, complementizers, prepositions, auxiliary verbs, etc. — in particular the monosyllabic versions of these — are not.” Additional candidates without footing of the first word are considered later in this section.

<table>
<thead>
<tr>
<th>/it as/</th>
<th>ONSET, MAX($\text{lex}$, $\omega$), WbyP</th>
<th>*µ/C</th>
<th>ALIGN</th>
<th>FtBIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>i t a s</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>i t a s</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>c.</td>
<td>i t a s</td>
<td></td>
<td>*</td>
<td>*!</td>
</tr>
</tbody>
</table>

Given the two rankings ONSET $\gg$ ALIGN and WbyP $\gg$ *µ/C, ambisyllabicity is untenable. Candidates (11b–c) exhibit ambisyllabicity of the resyllabified consonant. That is, [t] is shared by two syllables. In (b), [t] is moraic; in (c), it is not. Of these two candidates, (b) is incorrect.
on empirical grounds. As discussed in §1, resyllabified consonants do not contribute to weight, ruling out (b). Candidate (c), however, has the appropriate weight profile. Nevertheless, (c) is ruled out analytically. Because WbyP must dominate *μ/C, (b) is more harmonic than (c), and, as just mentioned, (b) cannot win. If ALIGN were ranked above WbyP or *μ/C, (b) would win, which cannot be allowed. In short, because (b) must lose, both (b–c) must lose. Resyllabification cannot induce ambisyllabicity in Latin. Candidates (11b–c) also violate CRISPEDGE (not shown), which in this case penalizes spreading a consonant across syllables (Ito and Mester 1999). The ranking of CRISPEDGE is irrelevant here, in that (a) wins regardless of how it is ranked.

Next, as tableau (12) illustrates, lexical it cannot cliticize to the following PWd. MAX(lex, ω) requires it to project its own PWd, ruling out (b). (For clarity, ω is indexed to a corresponding morphosyntactic word, if any.) The alternative indexation (c) satisfies MAX(lex, ω) by indexing it to the maximal ω, but is then eliminated by ALIGN, given that the nearest right edge of a PWd is two segments away, across as. Additional possible indexations can be similarly eliminated.

<table>
<thead>
<tr>
<th>/it as/</th>
<th>ONSET</th>
<th>MAX(lex, ω)</th>
<th>WbyP</th>
<th>*μ/C</th>
<th>ALIGN</th>
<th>FtBin</th>
</tr>
</thead>
<tbody>
<tr>
<td>f, ωit</td>
<td>f, ωit</td>
<td>σ</td>
<td>σ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>μ</td>
<td>μ μ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a.</td>
<td></td>
<td>i</td>
<td>t</td>
<td>a</td>
<td>s</td>
<td>*</td>
</tr>
<tr>
<td>ω</td>
<td></td>
<td>σ</td>
<td>σ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>μ</td>
<td>μ μ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td>i</td>
<td>t</td>
<td>a</td>
<td>s</td>
<td>*</td>
</tr>
<tr>
<td>ωit</td>
<td></td>
<td>σ</td>
<td>σ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>μ</td>
<td>μ μ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td></td>
<td>i</td>
<td>t</td>
<td>a</td>
<td>s</td>
<td>*</td>
</tr>
</tbody>
</table>
MAX(lex, ω) and ALIGN, as defined, ignore function words. (This is not to deny that function words might be stressed for other reasons, as discussed in §4.) In (13), FtBIN favors PWd adjunction, the correct outcome. Nevertheless, the analysis as it stands makes the unwanted prediction that a word like et would be stressed whenever preconsonantal, in which case the t would be parsed as a moraic coda, satisfying FtBIN. Various constraints could be added to militate against stressing function words in such contexts. One is DEP(ω, lex), which penalizes a PWd that does not correspond to a lexical word (Ito and Mester 2019). Any constraint militating against feet or stress (even indirectly, as with alignment) would also work, appropriately lowly ranked. Output-output correspondence in stress would be yet another possibility.

Finally, consider the fact that lengthening is a possible repair for degeneracy for isolated words but not for words affected by resyllabification. For example, /da/ “give” is lengthened to [(‘da:)], necessitating that FtBIN dominate DEP-μ. But given this ranking, what precludes /it as/ above from lengthening to *[‘(i:)‘(t as)]? Higher-ranking output-output correspondence, specifically, DEPOO-μ, penalizes lengthening if the word surfaces elsewhere without length. Thus, *[‘(i:)‘(t as)] is precluded by free-standing [(‘it)], but no output *[‘(da)] is found to stand in the way of [(‘da:)]. See Breiss (2021) on prosodic output-output correspondence more broadly.

3 Degeneracy avoidance in the Latin hexameter

As discussed in §1, in Virgil’s hexameter, C₀VC words freely occur prevocally if they are function words (41%), but not (except rarely) if they are lexical words (5%). This difference can be explained by stress. If C₀VC is unstressed, as with the function word ab “from” in (14a), resyllabification does not create a degenerate foot. If lexical monosyllables lost their stress prevocally, as with dat “gives” in (14b), Virgil would have no reason to avoid them prevocally. However,
lexical monosyllables retain their stress under resyllabification, as in (14c), permitting one to explain the rarity of that configuration in Virgil’s epics as reflecting a constraint against degenerate feet.

(14) a. a(b oː)rːiːs “from shores”
   b. *da(t oː)rːas “gives shores”
   c. (da)(t oː)rːas “gives shores”

This avoidance does not extend to stressed light syllables in general. Ryan (2017: 590, see also Mercado 2021) finds that stressed light syllables occur in Virgil’s Aeneid at nearly the same rate as in prose (indeed, in the pre-cadence, they are overrepresented relative to prose). Thus, Virgil’s avoidance of forms like (14c) is not due to the coincidence of lightness and stress; it is specific to degenerate feet. Moreover, lexical monosyllables with long vowels (C⁰V:C) are not avoided before vowels. They occur prevocally 57% of the time.⁸ Thus, Virgil’s avoidance of forms like (14c) is not due to the coincidence of monosyllabicity and stress.

Despite the avoidance of degenerate feet in Virgil’s (and certain others’) hexameter corpora, such avoidance is not characteristic of Latin in general, judging by prose, nor is it found in the verse of all authors. Figure 1 shows the rate at which C⁰VC is found prevocally for lexical vs. functional words for a several authors. As a sample of hexameter authors, I take (roughly chronologically) Horace, Virgil, Ovid, Lucan, Valerius Flaccus, Statius, Juvenal, and Silius Italicus. (Lexical C⁰VC is sufficiently infrequent that a large corpus is needed, which excludes certain other hexametrists, such as Ennius.) As a prose sample, I take Caesar, Cicero, Nepos, Livy, Seneca, and Quintilian. Approximate end dates are provided in the figure.⁹ The lists of prepositions and lexical words are the same as in §1. Prose corpora were split into lines by sentence punctuation.

As Figure 1 reveals, most of the hexameter corpora (with the exceptions of Horace and possibly Juvenal) exhibit strong avoidance of lexical C⁰VC before a vowel, whereas none of the prose corpora does. In particular, hexametrists Virgil, Lucan, Valerius Flaccus, Statius, and Silius Italicus nearly (though never categorically) forbid lexical C⁰VC in prevocalic position. Ovid likewise shows a significant difference between lexical and functional C⁰VC, though in his case, lexical C⁰VC remains frequent before vowels.

It therefore appears that avoidance of lexical C⁰VC before a vowel is specific to verse, reflecting the poetic grammars of certain hexametrists.¹⁰ These poetic grammars, I claim, possess a constraint forbidding a metrical subposition from containing a foot: *SUBPOS ⊃ FOOT. (Because degenerate feet always project to degenerate prosodic words in Latin, *SUBPOS ⊃ PWd would work equally well here.) This constraint belongs to a family of position size maxima (Hanson

---


⁹Specific works are as follows: Horace’s Sermones (Satires 1 and 2), Virgil’s Ecl. Eclogues, Georgics, and Aeneid, Ovid’s Metamorphoses, Lucan’s Phrasalia, Valerius Flaccus’ Argonautica, Statius’ Thebaid, Juvenal’s Satires, Silius Italicus’ Punic, Cicero’s oratories, Caesar’s war commentaries, Nepos’ De viris illustribus, Livy’s Ad urbe condita, Seneca’s Epistles morales ad Lucilium, Quaestiones naturales, and Dialogi, and Quintilian’s Institutiones and Declamationes maiores. All were accessed at the Latin Library (www.thelatinlibrary.com) on May 10, 2022.

¹⁰Verse vs. prose cannot be compared for any individual author here, as each author’s works fall entirely or nearly entirely on one side or the other. That said, given the range of authors and dates and the near consistency of the split between verse vs. prose, it is unlikely that the split can be attributed instead to dialect or chance.
Figure 1: Percentage of $C_0VC$ words that are prevocalic as opposed to preconsonantal, arranged by genre (labels at top), author (at bottom), and word type (grayscale fill). Within each genre, authors are arranged by rough *termini ad quem*, usually the end of the author’s life (to save space, “circa” is omitted). Throughout, error bars are 95% confidence intervals based on the binomial.

and Kiparsky 1996). By excluding degenerate feet from subpositions, the constraint effectively excludes degenerate feet from anywhere in the line, as such feet already could not occur elsewhere, being light. To account for the varying strength of $*\text{SUBPOS} \supset \text{FOOT}$ for different poets, the constraint could be weighted, as in Harmonic Grammar (Pater 2009).

This analysis based on $*\text{SUBPOS} \supset \text{FOOT}$ makes no predictions about degeneracy avoidance outside of verse. After all, subpositions are not found in prose. Moreover, because different poets possess different metrical grammars even for the same meter, $*\text{SUBPOS} \supset \text{FOOT}$ need not be active, or active to the same degree, for every poet and every type of verse. Individual poets, even when part of the same metrical tradition, vary in their permissiveness of licenses and their sensitivity to linguistic factors in ways that do not necessarily reflect differences of dialect (see e.g. Hayes et al. 2012 on the iambic pentameters of Shakespeare vs. Milton). In the present case, $*\text{SUBPOS} \supset \text{FOOT}$ would have different weights for, say, Ovid and Virgil. But the lack of evidence for $*\text{SUBPOS} \supset \text{FOOT}$ in Horace’s verse does not imply that Horace’s language lacks degenerate feet. It suggests only that his poetic grammar is insensitive to the category. Meters are not sensitive to every kind of linguistic structure available to the poet (consider the role of pitch accent in most quantitative meters).

4 Varying stress propensities of function words

The previous section identified five hexametrists who strongly (but not categorically) avoid lexical $C_0VC$ before a vowel, namely, Virgil, Lucan, Valerius Flaccus, Statius, and Silius Italicus. So far, I
have employed prepositions as representative function words and nouns and verbs as representative lexical words. In this section, I consider additional word types of shape C\textsubscript{0}VC, namely, adverbs, pronouns, and conjunctions. Here, adverbs include \textit{bis} “twice,” \textit{sat} “enough, well,” and \textit{ter} “thrice,” pronouns include \textit{id} “it,” \textit{quid} “what,” \textit{quis} “who,” and \textit{quod} “which,” and conjunctions include \textit{an} “whether,” \textit{at} “yet,” \textit{et} “and,” \textit{sed} “but,” \textit{tot} “so many,” \textit{ut} “that,” and \textit{uel/vel} “or.” As Figure 2 shows, these additional types are intermediate in behavior between lexical words and prepositions. They are significantly more avoided before vowels than prepositions, but significantly less than nouns or verbs.

Figure 3 shows the rate for each individual word, combining the works of the five authors into a single corpus. The rarer the word, the larger its error bar. Note that some words that appear to be outliers for their category, such as \textit{stet} “stand (subjunctive)” for lexical words, are sparsely attested ($n = 9$ in this case) and thus less reliable. Furthermore, categorization is not always straightforward. Lexical words here include nouns and verbs; adverbs are kept separate. Among adverbs, \textit{bis} “twice” and \textit{ter} “thrice” pattern more like lexical words, while \textit{sat} “enough, well” patterns more like a preposition.

On the present analysis, words’ varying rates of prevocalism reflect their varying propensities to be stressed. These stress propensities are general to the language — indeed, they are highly general across languages — and need not be stipulated by this analysis by, for instance, indexing *SUBPOS\begin{symbol}FOOT to individual words. The different treatment of lexical vs. functional words vis-à-vis stress and cliticization has been treated extensively (Selkirk 1996, 2011, Truckenbrodt 2007, Ito and Mester 2019). Meanwhile, pronouns and conjunctions vary in stress, as determined in part by pragmatic factors such as focus, givenness, topic, demonstrativity, and so forth (Lee et al. 2008). In sum, the range of prevocalic rates in Figure 3 reflects a cline of clisis, to use a phrase coined by Dieter Gunkel. The more likely a C\textsubscript{0}VC word is to be stressed, the more it will be avoided.
prevocally, as it is only stressed C₀VC words that incur violations of *SUBPOS⊃FOOT under resyllabification.

5 Degeneracy avoidance in the Ancient Greek hexameter

Like Latin, the Ancient Greek hexameter, as represented here by the Iliad and Odyssey, distinguishes degenerate feet from other light syllables. The basic template for the meter is the same as in (5), though Greek, unlike Latin, is a pitch accent language. I assume, following most modern analyses, that Ancient Greek word prosody — not just its poetic metrics — is organized around feet, which are invoked to motivate pitch accent placement, prosodic minimality, syllable timing, and other phenomena (Steriade 1988, Sauzet 1989, Golston 1990, Devine and Stephens 1994, Blumenfeld 2004, 2011, Sandell 2020; but cf. Steriade 2014). Here, all that is relevant is that monosyllables can be footed, normally so as lexical words, and sometimes as function words, just as in Latin and other languages.

As before, the present concern is monosyllables of the shape C₀VC. Short-voweled monosyllables exhibit three accentual patterns, namely, acute (high or rising), grave (lowered or canceled high; cf. Devine and Stephens 1994, Probert 2003, 2006), and none (certain clitics). An accented monosyllable is acute when at the end of an intonation group or hosting an enclitic; otherwise, it is grave. Regardless of accent and word type, resyllabification normally applies in C₀VC#(h)V throughout the line, as confirmed by scansion.

Figure 4 shows prevocalic rates of monosyllables in the combined Homeric corpus, aggregated by word type (lexical, pronoun, preposition, or conjunction). The corpus and word lists were
prepared as follows. The two Greek texts were downloaded from the Perseus Digital Library.\textsuperscript{11} Repeated lines were removed after the first instance. All words of shape \(C_0VC\) were taken as data, excluding those with long vowels (even when length is not orthographically apparent, as with some alphas, iotas, and upsilons), apocopic forms such as \textit{pot’} (for \textit{pote} “at some time”), and allomorphs confined to preconsonantal or prevocalic position (e.g. \textit{kád} for \textit{kata} “against” before \textit{d}). The emphatic enclitic \textit{per} was excluded because it does not fit well with any of the categories in the figure.\textsuperscript{12} Finally, any token of prevocalic \(C_0VC\) scanning as heavy was excluded.\textsuperscript{13}

After these exclusions, the following \(C_0VC\) words remain. Lexical words include \textit{bán} “go, mount,” \textit{dós} “give,” \textit{pʰán} “speak,” \textit{pʰtʰán} “come,” \textit{stán} “stand,” and \textit{tʰés} “put.”\textsuperscript{14} Because there is only one adverb, \textit{trís} “thrice,” and it patterns with the lexical words, the adverb is counted as lexical in this section. Pronouns include \textit{hén} “one,” \textit{min} “him/her/it,” \textit{són} “your,” \textit{sós} “your,” \textit{spʰin} “for them,” \textit{tís} “who,” \textit{tis} “someone,” and \textit{tón} “him.” Prepositions include \textit{en} “in,” \textit{es} “into,” \textit{ksún} “with,” \textit{prós} “towards,” and \textit{sún} “with.” Finally, conjunctions include \textit{`án} (modal particle), \textit{gár} “for, since,” \textit{ken} (modal particle), \textit{mén} “whereas,” and \textit{tár} “and so.”

As Figure 4 reveals, lexical \(C_0VC\) words, unlike the other word types, are almost categorically avoided before vowels. Their prevocalic incidence is 4.8\% (5 of 104 tokens), almost exactly matching Virgil’s rate of 5.0\% (6 of 119 tokens) in (10) and that of several other Latin hexametrists in Figure 1. The analysis is likewise the same. Due to \texttt{Max(\texttt{lex}, \texttt{\textomega})}, a lexical word, regardless of accent, must correspond with a PWd. If the lexical word is \(C_0VC\) and prevocalic, the PWd is

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Prevocalic rates for Greek monosyllables, aggregated by word type.}
\end{figure}

---

\textsuperscript{11}\url{www.perseus.tufts.edu}, accessed May 14, 2022. Further post-processing was facilitated by James Tauber’s \texttt{greek-accentuation} module for Python.

\textsuperscript{12}Even though it is not shown, \textit{per} agrees with the generalizations in this section, being prevocalic 65\% of the time when unaccented and 5\% of the time when acute.

\textsuperscript{13}This includes any prevocalic \(C_0VC\) in the (normally heavy) line-initial position, the site of some irregularity (West 1982). It also includes a number of tokens preceding a (historical) digamma (*\textit{w*}), which is not indicated in orthography but which usually blocks resyllabification. For example, \textit{hói “they”} scans as if it were \textit{hwoi}. Using a list of common digamma-initial words from Monro (1891), any bigram with a digamma-initial second member was excluded. Additionally, all remaining tokens of prevocalic monosyllables except (as accented) \textit{`án}, \textit{en}, \textit{gár}, \textit{ken}, \textit{mén}, \textit{min}, and \textit{tís} were checked by hand. Among the 444 tokens scanned, 1.6\% involved digammas not already captured. The error rate for the remaining words is likely similar.

\textsuperscript{14}The \textit{n}-final forms are neuter singular aorist active participles. The \textit{s}-final forms are second-person singular aorist active imperatives.
rendered light by resyllabification, as in (15). The resulting syllable, being light, cannot occupy any position in the meter that requires a heavy syllable. But it also cannot occupy a light-requiring subposition, given *SUBPOS\supset FOOT (weighted as necessary to permit exceptions). It follows that degenerate feet are disfavored throughout the hexameter.

(15) (dō)(s apophēmenon) “grant that he meet his demise”

The presence of pitch accent in Greek raises a possible confound for this analysis that I now address. Hypothetically, one could imagine that lexical monosyllables might be more likely to be acute than other monosyllables, and that it might be their acuteness rather than their lexicality driving their underrepresentation before vowels. Figure 5 breaks down word type by accent type, showing that accent type does matter, but not in a way that undermines the generalization just posited. To wit, two generalizations are now evident: First, lexical C₀VC is avoided prevocally regardless of accent (acute or grave), as already stated. Second, acute C₀VC is avoided prevocally regardless of word type. Even C₀VC function words are nearly unattested before vowels when they are acute.

Analyzing this second generalization is beyond the scope of this article, but possibly represents another way in which degenerate words are avoided in subpositions. Recall that monosyllables are acute in two contexts, namely, when final in an intonation group and when hosting an enclitic.¹⁵ Both contexts have been analyzed elsewhere as being associated with prominence. Ito and Mester (2019), for one, treat the non-reducibility of function words when final in the phonological phrase in English. Likewise, clitics are pre-stressing (requiring their host to be footed) in some languages (Mester 1994). That said, I leave these matters in Ancient Greek to future research. What is important here is that the lexicality gradient persists even when accent is controlled, as with grave accent in Figure 5.

Finally, as in Latin, beyond lexical categories, individual Greek C₀VC words vary in their prevocalic propensities, presumably reflecting varying propensities to be stressed. Figure 6 shows the rates for individual monosyllables in Greek. Note that each word has a separate entry for each accent pattern that it takes. For example, tís “who,” which is rarely prevocalic, is separate from tis “someone,” which is often prevocalic.

6 Discussion

A degenerate foot, despite being a kind of light syllable, is penalized in light-requiring positions in meter, being nearly banned by some poets. The present analysis does not maintain that the poets avoid degenerate feet simply because such feet are marked. (If that were the case, there would be no explanation for why the poets avoid one type of marked structure but not numerous others, nor would there be an explanation for the confinement of the avoidance to verse.) Rather, the poets avoid degenerate feet because they exceed the tolerance of a light-requiring metrical position. That is, a light-requiring position (subposition in this case) cannot dominate a prosodic category higher than the syllable.

¹⁵ These contexts need not be expressed disjunctively. Assuming that an enclitic projects to a recursive prosodic word, a non-disjunctive generalization is that an acute ultima of a maximal prosodic word becomes grave when non-final in the intonation group. Ultimas before enclitics are not affected because they are not final in their maximal prosodic words.
Figure 5: Prevocalic rates for Greek monosyllables, grouped by accent type (top) and word type (bottom).

Figure 6: Prevocalic rates for Greek monosyllables, showing individual words. The labels at bottom also indicate the prevocalic and total frequencies of each word.
Such avoidance is phonetically natural in the sense that a degenerate foot is presumably the longest type of light syllable in the language. At least three factors potentially conspire to make degenerate feet longer than other light syllables in Latin and Greek. First, on the present analysis, a degenerate foot arises only in a monosyllable, and monosyllables tend to be longer cross-linguistically than syllables that are part of longer words and therefore subject to polysyllabic shortening (Lehiste 1970, Mori 2002, 2014). Second, degenerate feet are stressed, and stressed syllables tend to be longer than their unstressed counterparts, as evidenced in Latin by reductive sound changes (Weiss 2020). Finally, degenerate feet, being always word-final on this analysis, have the potential to be phrase-final and thus undergo final lengthening (cf. Delattre 1966, Lindblom 1968, Wightman et al. 1992), while non-final light syllables lack this potential.

With this phonetic rationale in mind, this section briefly addresses the viability of alternative possible analyses of the avoidance of stressed CV.CV#V in verse that do not depend on the foot. First, as mentioned in §3, *SUBPOS\sup{PWD} would work just as well as *SUBPOS\sup{FOOT} here, since degenerate feet are always coextensive with degenerate prosodic words on this analysis. Thus, if one rejects the foot but not the prosodic word, an alternative is available that trivially modifies the present analysis.

Consider next a purely grid-based approach, without any bracketing for feet or prosodic words. As (16) illustrates, the avoidance of stressed CV.CV#V does not reflect the avoidance of a particular rhythmic configuration, at least not one expressed by the standard grid. For example, avoided type (16a) *dat amīca* has the same rhythmic profile as unavoidable type (16b) *data porta*. To augment the excessive crudeness of the traditional grid, one might pursue one of two strategies, invoking either monosyllabicity or phonetic duration.

(16) a. *dat amīca*  
“friend gives”  
\[
\begin{array}{cccc}
\times & \times & \times & \times \\
\times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
\text{dat} # a \ mi : \ ka
\end{array}
\]

b. *data porta*  
“door given”  
\[
\begin{array}{cccc}
\times & \times & \times & \times \\
\times & \times & \times & \times \\
\times & \times & \times & \times & \times \\
\text{da} . t \ a \ # \ por \ ta
\end{array}
\]

First, consider a constraint like *SUBPOS\sup{STRESSED\_MONO} “a subposition cannot contain a stressed monosyllable.” The first issue with this constraint is that it defeats the intent of the purely grid-based strawman by invoking what can only be construed as a type of prosodic word, namely, the monosyllable. As just discussed, if one allows reference to the prosodic word, then a constraint like *SUBPOS\sup{PWD} suffices, which has the effect of excluding degenerate prosodic words from the hexameter without invoking monosyllabicity. Another objection to an analysis in terms of *SUBPOS\sup{STRESSED\_MONO} is that it stipulates the description, giving up a fair amount of restrictiveness in so doing. For instance, the constraint *SUBPOS\sup{STRESSED\_MONO} implies the simpler constraint *SUBPOS\sup{MONO}, banning all monosyllables from subpositions. I am not aware of such a meter.\(^{16}\) The *SUBPOS\sup{FOOT} analysis, in which a position type exhibits a prosodically defined maximum (Hanson and Kiparsky 1996), avoids such overgeneration. “Monosyllable,” not being a prosodic category, is not available as a predicate.

A final possible line of analysis that eschews the foot is to refer directly to (normalized) pho-

\(^{16}\)The avoidance of monosyllables in line-final position is not a function of position type.
netic duration, defining duration ranges for metrical positions (cf. Flemming 2001, Ryan 2014). On such an approach, stressed monosyllables would exceed the durational tolerance of a light-requiring position. Restrictiveness, once again, is a concern. For example, if the cutoff for light positions can be drawn at whichever point is necessary to preclude stressed monosyllables, what prevents other cutoffs, such as excluding stressed monosyllabic $C_0V$, but only if the vowel is low? Another concern is the abstract computation of duration, that is, normalization. For example, one would presumably not want positional tolerances to vary as a function of speech rate, as raw duration would predict. Moreover, categories tend to overlap. For instance, the shortest heavies are often shorter than the longest lights, even when considering mean durations (Hayes 1999, Gordon 2002). Thus, a structure-free approach appears to be a nonstarter. That said, poetic practice can be sensitive to fine-grained phonetic detail (Ryan 2011), leaving open the question of how best to integrate phonetic and phonological phenomena in metrics.

7 Conclusion

In Latin and Ancient Greek, lexical monosyllables must remain prosodic words when they undergo resyllabification, in effect becoming degenerate feet. Verse evidence confirms that such monosyllables retain their stress even while becoming light. Hexametrists in both the Latin and Greek traditions avoided degenerate feet in their verse due, I maintain, to a constraint against a subposition dominating a foot. Meanwhile, prevocalic lexical $C_0VC$ was permitted freely outside of verse, revealing that the languages in general tolerated degeneracy. Beyond lexical words, as this article recognizes, function words vary in their propensities to be stressed. The more likely a $C_0VC$ function word is to be stressed, the more it is avoided before a vowel in the hexameter, reflecting the same degeneracy effect observed more rigidly for lexical words. With this new evidence for degenerate feet being induced by resyllabification in languages that otherwise do not permit degeneracy, the degenerate foot receives additional support, being more widely distributed in the typology than previously acknowledged.

References


