# The syllable*

Péter Szigetvári

Syllables are one of the few phonological notions that the average layman knows about. Others include sounds, consonants and vowels.[1] Syllables belong to the prosodic domain of sound structure, together with feet: both categories are related to the rhythm of speech. While feet are the manifestation of the rhythmic alternation of stressed and unstressed portions, syllables represent the alternation of high-sonority (syllabic) and low-sonority (nonsyllabic) portions, or vowels and consonants, as they are usually referred to. These two waves, the rhythm wave and the sonority wave, carry the speech signal. Their interference is constructive, i.e., a rhythmic peak always coincides with a sonority peak, so the two waves never cancel out.

Since, unlike morphemes, syllables are not associated with meaning, semantic considerations do not enable the speaker to discern where one syllable ends and the next one begins. In fact, it is not even obvious if phonological theory has to include an entity that corresponds to syllables. It is well-known that syllables are claimed to be unnecessary by Chomsky & Halle (1968) in what came to be a cornerstone of modern phonological theory, the SPE (*The Sound Pattern of English*). Section 1 shows that although technically viable, this solution misses a generalization, the recurrent environment that can be identified as the end of the syllable, which can only be expressed disjunctively in the classical version of the generative theory. On the other hand, the syllable as a unit of phonological representations does not appear to be absolutely necessary. Section 2 examines whether syllable structure can be derived from the order of sound strings, or at least some part of it must be given lexically. The impossibility of equating word and syllable boundaries is

* I am indebted to two anonymous reviewers, András Cser, Ádám Nádasdy, and Péter Siptár for comments on this paper. Errors are my responsibility.

[1] Think of Regina Spektor's Consequence of sounds, for example.

demonstrated. This calls for an exclusively sonority based syllabification, which also turns out to be problematic. The conclusion is that syllabicity must be lexically marked.

Arguments for the possible internal organization of the syllable are looked at in section 3. They are based on syllable weight (section 3.1), phonotactic constraints (section 3.2), and closed syllable shortening (section 3.3). None of these are capable of clearly distinguishing between the CV+C and the C+VC basic split within a syllable. Section 4 introduces the option of having 'syllables' without a pronounced vowel, a move that reshuffles most of what has been thought of about syllables. (Interestingly, the first written record of this analysis comes from the middle of the 18th century.) This idea is elaborated in section 5, showing that a model that involves only relations between pairs of segments and dispenses with the prosodic hierarchy above them has certain advantages over traditional models of syllabic structure in representing consonant clusters (section 5.1), long vowels (section 5.2), the status of intervocalic consonants (section 5.3), and syllable weight (section 5.4).

## 1 Why have syllables?

Early generative phonological descriptions avoid explicit reference to the syllable.[2] They deny not only the level of the syllable in the prosodic hierarchy, but also its boundaries, although representations are crowded with boundary symbols — albeit mostly grammatical ones. The SPE, for example, proposes the rule in (1) for French elision▷ and liaison▷(Chomsky & Halle 1968: 353). ◁ elision ◁ liaison

(1) *SPE elision and liaison rule*

$$\begin{bmatrix} -\alpha\,\text{vocalic} \\ \alpha\,\text{consonantal} \end{bmatrix} \rightarrow \emptyset \;\Big/\; \underline{\quad}\;\#\,[\alpha\,\text{consonantal}]$$

In its own clumsy way the rule aims at retaining the consonant–vowel alternation across a word boundary by deleting a consonant

[2] McCawley (1965) is a notable exception.

before a consonant and a vowel before a vowel. Note that the rule in (1) deletes a vowel, and does not delete a consonant, before a glide (e.g., *l'oiseau* [lwazo] 'the bird' and *petit oiseau* [pətitwazo] 'little bird'). But since glide-initial words labelled [+foreign] by Chomsky & Halle pattern with consonant-initial words (e.g., *le yogi, le watt*), the feature [syllabic] is tentatively introduced to replace the original [vocalic]. The rule in (2) takes care of elision and liaison in foreign words (1968: 355).

(2) *SPE elision and liaison rule for foreign words*

$$\begin{bmatrix} -\alpha\,\text{syllabic} \\ \alpha\,\text{consonantal} \end{bmatrix} \rightarrow \emptyset \;\Big/\; \underline{\quad}\,\#\,\begin{bmatrix} -\alpha\,\text{syllabic} \\ +\,\text{foreign} \end{bmatrix}$$

We see in (2) that Chomsky & Halle are pressed to posit a feature [±syllabic] in a framework that otherwise denies the necessity of including this notion in its formalism.

To explain the divergent behaviour of obstruent+glide/r consonant clusters with respect to stress placement rules, the definition of a weak cluster▹(the SPE equivalent of a light syllable) in (3) is arrived at (1968: 83). ◃ weak cluster

(3) *The weak cluster in SPE*

$$\begin{bmatrix} -\text{tense} \\ \text{V} \end{bmatrix} \text{C}_0^1 \begin{bmatrix} \alpha\,\text{vocalic} \\ \alpha\,\text{consonantal} \\ -\text{anterior} \end{bmatrix}_0$$

The formula in (3) includes the feature [−anterior] to exclude [l] (while including [j w r]), since in their analysis words like *armadíllo, vanílla, umbrélla* contain a geminate /ll/, hence a strong cluster▹, which as a result is stressed. However, words like *díscipline, pánoply,* or in Chomsky & Halle's analysis *míracle, clávicle* etc. (1968: 197) show that stop+l sequences do create a weak cluster, since we find antepenultimate stress in these words, which is expected only if the ◃ strong cluster penultimate vowel is followed by a single consonant or a weak cluster (e.g., *cítizen, mímicry*). Thus, besides being extremely unnatural, the definition of the weak cluster is empirically inadequate.

Kahn (1976: 22ff) notes that the environment $\_\_\{ {C \atop \#} \}$ is recurrent in many phonological rules across languages, it is here, for example, that [l] is velarized and [r] is deleted in Standard British English. The attempts at making consonants and the word boundary a natural class to avoid the disjunction turned out to be unsuccessful. Chomsky & Halle seem to be missing a generalization here, the syllable boundary.

Subsequent research has adopted the notion of the syllable in phonological theory collecting a set of arguments for its necessity. Phonotactic constraints are often referred to for justifying syllables. But, as Blevins admits, 'surprisingly, there are few if any feature co-occurrence constraints which appear to take the syllable as their domain' (1995: 237, fn. 8). In fact, as I will argue in section 3.2, phonotactic constraints hold primarily between adjacent consonants (and adjacent vowels, that is, diphthongs). Crucially, interdependent consonants may or may not be tautosyllabic, it is their adjacency that matters. A tautosyllabic onset and coda[3] do not have any effect on each other, heterosyllabic coda+onset clusters, however, do. Therefore the syllable is not a relevant domain for formulating phonotactic constraints.

◁ phonotactic constraints

Another argument for the unity of the syllable is that stress anchors not on vowels, but on whole syllables (Hayes 1995: 49f). Halle & Vergnaud claim that vowels are the stress-bearing elements: in some languages stress can fall on either mora of a long vowel (1987: 49, 61, 193). If this is a possibility, the syllable is not indispensable for locating stress. But even if it were the case that only the first mora of a syllable could bear stress, one could still argue that the anchor for stress is not the syllable but the nucleus or the rhyme in it.

◁ stress-bearing element

[3] In this chapter, I use terms like 'onset', 'coda', or 'syllable' in a descriptive sense, as it will become clear, their theoretical status is not firm.

native speaker intuitions

As has already been mentioned, native speakers have intuitions about syllables in their language. Harris & Gussmann (1998) argue that these intuitions are the result of linguistic traditions. They examine the case of word-final consonants, which, according to what they refer to as the 'western' or Graeco-Roman tradition, belong to the same syllable as the preceding segment(s). According to a more ancient 'eastern' tradition, such a consonant is the onset of a syllable that lacks an audible nucleus. Native speaker intuitions about syllables — at least some of them — are thus not based on unconscious linguistic analysis, but on traditions about language learnt at school. They should accordingly be treated with suspicion.

Before one should hastily conclude that the notion of the syllable is useless, it is worthwhile clarifying what it means 'to have' syllables in one's theory. The periodic alternation of low- and high-sonority portions in sound sequences is an empirical fact of sound-based human languages (noticed by Sievers 1881), this fact must be accounted for. If the way a particular theory accounts for the sonority wave in speech is called 'syllable', then any theory is bound to have syllables. If, however, we restrict the use of this term to an explicit node in the prosodic hierarchy, then theories will vary to a large extent in having or not having syllables. Accordingly, the only meaningful way in which two theories can be distinct in one having, the other one lacking syllables is having an explicit bit in the representation that is labelled 'syllable', or 'rhyme', or at least having a 'syllable boundary'. But even if a theory does not recognize the syllable explicitly, the term can be used in a descriptive way to denote a sonority cycle of the speech signal.

## 2 Underlying or derived?

Blevins lists three reasons for assuming that syllable structure is not underlying, but derived (1995: 221): (i) the scarcity of minimal pairs distinguished only by syllabification, (ii) syllabicity alternations (e.g., *opin*[iə]*n*∼*opin*[jə]*n*); and (iii) morphemes often cannot be parsed according to the syllabification rules of the language. The first two reasons are closely related, and are primarily a result of the ambiguous status of glides. This will be discussed in section 2.3. Let us first see if reason (iii) provides a solid foundation for derived syllabification.

The ill-formedness of the syllable structure of morphemes in the lexicon is trivially true for suffixes consisting only of consonants. The English past tense suffix *-d,* for example, cannot be syllabified to create a well-formed word. This is not surprising: it *is* not a well-formed word. However, it is not only affixes but also stems that resist exhaustive syllabification. A word like *hymnal* [hɪmnəl] suggests that *hymn* [hɪm] ends in the cluster [mn] in the lexicon. Since this cluster does not occur syllable finally, it seems to be a plausible assumption that lexically it is represented as an unsyllabified sound string //hɪmn//, which surfaces as [hɪm] after it turns out that the final [n] cannot be incorporated by syllabification and hence is deleted by stray erasure▷(Steriade 1982).[4] When suffixed with *-al,* the final [n] is rescued because it becomes the onset of the second syllable, therefore it is pronounced.

◁ stray erasure

Such facts argue against underlying syllabification only if we accept two further assumptions: namely, that a lexical form must be exhaustively syllabified and that the stem of *hymnal* is a lexical form. Neither assumption is self-evident. The morpheme *hymn* could lexically contain a fully syllabified part [hɪm], followed by an unsyllabified [n]. More importantly, it is not obvious that //hɪmn// is an entry of the lexicon. Kaye demonstrates that forms involving nonanalytical morphology▷(of which *-al* suffixation is an example) are invisible to phonology (1995: 308f). This means that *hymnal* is a lexical entry in its entirety, creating a single phonological domain, therefore phonology is ignorant of its internal morphological structure. In the case of analytical affixation,▷the situation is different: the stem of *hymning* [hɪmɪŋ] does undergo phonological processing. As a result it surfaces in the same form as the unsuffixed stem *hymn.* What this means is that there is a lexical entry //hɪm// (without a final [n]), and another entry //hɪmnəl//, the two are not

◁ nonanalytical morphology

◁ analytical morphology

[4] The alternative solution is to insert a vowel to save such a consonant. This happens in *rhythm,* where //rɪðm// (cf. *rhythmic* [rɪðmɪk]) surfaces as [rɪðəm].

related by phonological rules. According to this view, lexical entries are free forms (complete words), hence their syllabification can be lexical. As for affixes, nonanalytical ones (like *-al*) do not occur independently in the lexicon, while analytical ones (like *-ing* or *-d*) are not phonological domains in themselves, they create a domain together with the stem they attach to.[5]

If the syllabification of a string is predictable from the phonetic properties of the segments it is made up of, the question whether syllable structure is lexically given or derived is only relevant if there are phonological rules that apply lexically prior to syllabification took place. Without lexical rules, the question loses its significance. The point of having syllabification algorithms is then to check the phonotactic validity of segment strings.

## 2.1 Syllabification with reference to word edges

Kahn's (1976) syllabification algorithm consist of a set of rules that identify the syllabic segments in a string and then stack nonsyllabic segments to it, first to its left, then to its right (these latter rules are referred to as rule IIa and IIb, respectively). The order is important: if an intervocalic consonant (cluster) can be syllabified to both the preceding and the following vowel, the latter choice is preferred: resulting in onset maximalization. Thus *Lisa* is split as [liː.sə] (although it could be [liːs.ə]) and *supra* as [suː.prə] (although it could be [suːp.rə], but *hangar* as [hæŋ.ə] and *anger* as [æŋ.gə] (here no other option is available). Kahn claims that the set of permissible syllable-initial (and syllable-final) consonant clusters is part of rules IIa and IIb (1976: 45). Actually, permissible clusters at the edges of the syllable are exactly those that we observe at the edges

◁ onset maximalization

[5] The distinction between analytical and nonanalytical affixation is roughly equivalent to that between level-1/lexical and level-2/postlexical affixation in lexical phonology. The crucial difference is that lexical affixation is denied any phonological relevance. In modern terminology, there is no spellout at the lexical level (Scheer forthcoming).

of the word, the reason being that words that cannot be exhaustively syllabified by rules IIa and IIb are ungrammatical. Thus, although the stacking operation of rule IIa is not controlled by reference to the set of word-initial consonant clusters but by a list which is part of the rule, the two sets are equivalent, so in effect a syllable-initial cluster is possible if that cluster occurs at the beginning of words in the given language, and the same holds for syllable- and word-final clusters.

The effect of this algorithm could be exemplified by the parsing of *anxiety*: in the classical generative model its morphological connection to *anxious* [æŋkʃəs] suggests an underlying [-ŋgz-] in *anxiety.* Since [-ŋg] does not occur syllable/word finally and [gz-] does not occur syllable/word initially, the form [æŋgzaɪəti] should be unparsable and hence ungrammatical. Wells (1990), nevertheless, does list it as a possibility. Lowenstamm mentions similar problems in Finnish, where three-member internal clusters occur, but two-member clusters are not found at either edge of a word (1981: 601). Ancient Greek has -lp- within words (e.g., [elpis] 'hope'), but neither initial *#lp-, nor final *-l# or *-lp# is found, thus there seems to be no way to syllabify this word, if we want to maintain that the phonotactics of syllable and word edges is the same.

Kahn's predictions are defied at another point by Ancient Greek. His algorithm takes consonants one by one to check whether the given string is a possible cluster or not. Thus if $VC_i$ is a nonexistent string, $VC_iC_j$ must also be impossible, since the algorithm has no chance to examine the longer cluster if it stops at the shorter one. Now word finally, as we have seen, *-l# is not possible, but -ls# is (e.g., [hals] 'salt'), albeit marginally. The situation is the same with *-p# vs. -ps# (e.g., [gyːps] 'vulture', [ɔːps] 'eye', etc.), with a larger set of examples. Longer clusters also exhibit this unexpected pattern: *-ŋk# vs. -ŋks# (e.g., [laryŋks] 'larynx', [sp$^{h}$yŋks] 'sphynx').

## 2.2 Lowenstamm's universal syllable

Lowenstamm (1981) shows that Kahn's algorithm is not only empirically but also theoretically faulty, since it cannot syllabify strings that are not encountered on the surface.[6] He claims that in [gɪdɔrɪm] (the plural form of Yiddish [gedɪr] 'limitation') the first vowel is inserted by a rule. Epenthesis thus takes the input [gdɔrɪm] and yields [gɪdɔrɪm]. Since Yiddish does not exhibit the cluster [gd-] word initially, [g] must be left unsyllabified by Kahn's rule IIa. But if it is not incorporated into any syllable, the epenthesis rule cannot apply, in fact, the whole form is discarded as ungrammatical. The rationale of the epenthesis rule is exactly to correct the ill-formed word-initial cluster, but it has no chance to do so in Kahn's framework.

Lowenstamm proposes a return to the notion of sonority in assigning segments to syllables. Following Jespersen (1920) and Grammont (1933), he defines a 'universal syllable▷ as in (4).[7] ◁ universal syllable

(4) *The Universal Syllable*

In a string of segments, a syllable is a maximal substring such that

a. (i) no segment is lower on the hierarchy than both its immediate neighbours
   (ii) no two segments of equal ranking on the hierarchy are adjacent
b. the onset is maximal within the limits of (a)

The hierarchy referred to in (4) is, of course, the sonority hierarchy. The effect of (4ai) is that within a syllable sonority first rises to a peak, then it falls away from the peak. Both the rising phase and the falling phase may be missing, but a fall in sonority may never be followed by a rise within any syllable. This is the sonority sequencing principle▷ The sonority of segments thus seems to play an indispensable role in defining possible and impossible syllables. Let us examine sonority. ◁ sonority sequencing principle

[6] Kahn clearly states that his theory is about syllables 'on the phonetic level' (1976: 20), Lowenstamm talks about 'underlying' syllables. What follows is that superficial syllabification cannot be achieved by an algorithm.

[7] (4a) is equivalent to Selkirk's sonority sequencing generalization (1984: 116).

## 2.3 Sonority

The sonority hierarchy is a ranklist of segment types — in some analysts' view of individual segments (e.g., Ladefoged 1993) — from least to most sonorous. A phonetic definition of sonority and whether the sonority of segments is a primitive of phonological theory (e.g., Kiparsky 1979) or derived from their phonetic properties (e.g., Clements 1990, Harris 1990) need not concern us here.

◁ sonority hierarchies

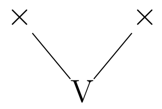
For some languages a very crude sonority hierarchy,▷containing two steps, will do: C→V (where the arrow points from minimal to maximal sonority). These languages will have neither consonant, nor vowel clusters (i.e., long vowels or diphthongs), but a strict alternation of consonants and vowels, i.e.,#(C)VCV… C(V)#. Some languages that allow a maximum of two consonants to occur next to each other may also be analysable with this simple scale, since in such languages (e.g., Japanese or Manam) the boundary of the syllable will always be between the consonants, as ruled by (4aii).[8]

An interesting situation arises in systems that contain long vowels. If a syllable with a long vowel is marked as CVV, this appears to violate (4aii): two segments of equal ranking are adjacent within a syllable.[9] The problem is only apparent, and lies in our linear–nonlinear hybrid rendering of the long vowel. In a genuine linear framework a long vowel is just another simple feature matrix (i.e., a single segment) with the feature [+long] included. In a nonlinear representation a long vowel looks like (5), where V stands for the melody of any vowel.

[8] There are no languages with syllable-initial consonant clusters, but no closed syllables. Kaye & Lowenstamm (1981) discuss implicational relations between syllable types, Blevins (1995) discusses syllable typology, and Piggott (1999) consonant cluster typology.

[9] Note that with consonants this problem does not occur: intervocalic geminate consonants are heterosyllabic. As will be argued below, word-final consonant clusters do not belong to the same syllable.

(5) *The nonlinear representation of a long vowel*




That is, the graphical reperesentation 'VV' for a long vowel is the result of merging two tiers, the skeletal and the melodic tier. If we are to keep using sonority sequencing, a long vowel is to be taken as a single unit.

To get diphthongs under the control of (4) the sonority hierarchy has to be refined: the glide part of the diphthong has to be assigned a sonority value smaller than the vowel part.[10] The resulting scale is this: O→S→V (where O means obstruent, S means sonorant consonants, this scale is adopted by Zec 1995).

This three-term sonority scale is enough for controlling sonority sequencing in an overwhelming majority of cases. Apart from [s], two obstruents or two sonorant consonants rarely (if ever) occur together on the same side of the vowel within a syllable (i.e., *OOV, *SSV, *VOO, *VSS), consequently obstruents or sonorant consonants do not have to be sequenced among themselves, and so they do not have to be ranked. Departures from this setting occur predominantly at word edges, but as we are going to see in section 4, there is reason to reject the automatical tautosyllabic analysis of any word-initial or word-final string of consonants. If such a view can be supported, then words like *film, act, sky,* etc. do not require ranking sonorants or consonants in the sonority scale. This simple scale becomes inadequate for systems where sonorant consonants can form the sonority peak in the syllable: for example, the first two consonants of Czech *mlha* 'fog' would be split by a syllable boundary, if all sonorants were of the same rank. In reality, the syllabification is [ml̩.ɦa], with syllabic [l̩]. The resulting scale is the following:

[10] Provided that these 'nuclear' glides are not categorized as C, in which case they *are* less sonorous than the preceding vowel.

O→N→S→V (where N stands for nasal, S for nonnasal sonorant consonants).

Another case where the ranking of sonorants seems necessary is for words like English *wild* and *kind* where a diphthong occurs in a closed syllable. Here the glide and the liquid/nasal have to be ordered, such that the glide is higher on the scale: O→N→L→G→V (where L means liquids and G glides; this scale is advocated by Clements (1990)).[11]

◃ syllabic consonant formation in English

The patterns of syllabic consonant formation in English▹also call for the splitting of liquids and nasals, but on the other hand, it questions the previous split of glides and other sonorant consonants. Let us first see the basic pattern in (6). (The judgements are those of Wells (1990).)

[11] Although there is a tradition of positing a ranking among obstruents along the voiceless→voiced and/or plosive→fricative dimensions (e.g., Anderson & Ewen 1987), this is to explain lenition and is unnecessary for the purposes of sonority sequencing in syllables, as Clements (1990) argues. Fricative+plosive clusters—especially with [s]—may occur both initially and finally (e.g., *stop/post*), or in reverse order (e.g., *ask/axe*). In English this is clearly not controlled by sonority sequencing. In Dell & Elmedlaoui's (1985) analysis of Imdlawn Tashlhiyt Berber, sonority distinctions within obstruents are required to derive the syllabicity of consonants. Guerssel (1990) and Coleman (1999), however, question the existence of syllabic obstuents and claim that reduced vowels occur next to these.

(6) *Syllabic consonants in English: the basic pattern*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| a. | little | lɪtəl | ∼ -tl̩ | b. | colonel | kɜːnəl | ∼ -nl̩ |
| | people | piːpəl | ∼ -pl̩ | | camel | kaməl | ∼ -ml̩ |
| | hassle | hasəl | ∼ -sl̩ | | plenary | pliːnəri | ∼ -nr̩- |
| | awful | ɔːfəl | ∼ -fl̩ | | camera | kamərə | ∼ -mr̩- |
| | cotton | kɒtən | ∼ -tn̩ | | salary | saləri | ∼ -lr̩- |
| | special | spɛʃəl | ∼ -ʃl̩ | c. | melon | mɛlən | ∼ *-ln̩ |
| | Belgian | bɛlʤən | ∼ -ʤn̩ | | column | kɒləm | ∼ *-lm̩ |
| | fathom | faðəm | ∼ -ðm̩ | | canon | kanən | ∼ *-nn̩ |
| | Benjamin | bɛnʤəmɪn | ∼ -ʤm̩- | | minimum | mɪnɪməm | ∼ *-mm̩ |
| | luxury | lʌkʃəri | ∼ -ʃr̩- | | lemon | lɛmən | ∼ *-mn̩ |
| | peppery | pɛpəri | ∼ -pr̩- | | venom | vɛnəm | ∼ *-nm̩ |

The data in (6a) aim to illustrate that a sonorant consonant may be syllabic after any obstruent.[12] The obvious and common explanation for this fact is that obstruents are less sonorous than sonorants, accordingly the syllabic sonorant remains the sonority peak after the loss of the vowel. As (6b) and (6c) show, syllabic consonant formation is more restricted after sonorant consonants. The data suggest the following ranking within sonorants: m, n→l→r. Syllabic glides (i.e., high vowels) may occur after any sonorant consonant, that is, they should be located at the top of the sonorant subscale: m, n→l→r→i/j, u/w. The data presented in (7) however partially reverse the ranking.

(7) *Syllabic consonants in English after glides*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| a. | barrel | barəl | ∼ -rl̩ | b. | narwhal | nɑːwəl | ∼ -wl̩ |
| | barren | barən | ∼ -rn̩ | | loyal | lɔɪjəl | ∼ -jl̩ |
| | quorum | kwɔːrəm | ∼ -rm̩ | | tower | taʊwər | ∼ -wr̩ |
| | terrorist | tɛrərɪst | ∼ -rr̩- | | lawyer | lɒːjər | ∼ -jr̩ |

[12] Exceptions (like *album* *[-bm̩]) are probably not related to sonority. Another constraint allows syllabic [r̩]—and nonsyllabic [r] for that matter—only prevocalically.

These data show that syllabic [l̩] and syllabic nasals behave as if they were higher up the sonority scale than [r], (7a), and the glides [j] and [w],▷(7b).[13] This means that the syllabic and nonsyllabic versions of high vowels/glides occupy separate positions on the sonority scale: the nonsyllabic versions are below all sonorant consonants, the syllabic versions are above them, together with vowels. If so, then either the high glides [j] and [w] and — what invites to be called the low glide — [r] and their syllabic (i.e., vocalic) counterparts are different entities, or syllable structure cannot be derived, but is lexically given. Obviously, the syllabicity of these segments cannot be derived by reference to their sonority, if we want to derive their sonority by reference to their syllabicity. (8) depicts this lexical marking: V stands for the nucleus or peak of the syllable, whatever is used to mark it, and C for the syllable margin.

◁ glides in English

(8) *Syllabic and nonsyllabic 'glides'*

| | | | | | |
|---|---|---|---|---|---|
| a. [i] | b. [j] | c. [u] | d. [w] | e. [ɚ] | f. [r] |
| V | C | V | C | V | C |
| \| | \| | \| | \| | \| | \| |
| i | i | u | u | r | r |

(8a), (8c), and (8e) represent the versions of these segments occurring in the syllable peak, (8b), (8d), and (8f) represent their syllable marginal occurrence.

Accordingly, at least syllabicity must be lexically given, and cannot be derived. If syllabicity is lexically marked, the three-level

[13] The last two items in (7b) are given in a rhotic pronunciation, which has syllabic [r̩] in this position.

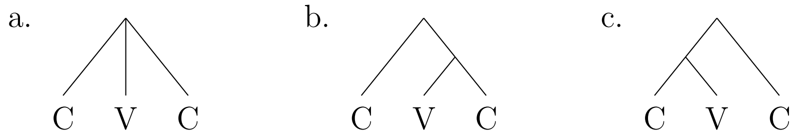
sonority hierarchy, O→S→V, suffices for practically all syllabification purposes.[14]

## 3 Internal structure

We will now consider what the possible internal organization of the syllable▷ could look like. (Yip 2003 gives a similarly sceptical account from a different perspective.) The first important issue to tackle is whether syllable structure is universal or language specific. The question is not whether any two syllables will have identical structure across languages: this would mean that all languages had to follow the one with the most complex structure.[15] The question is whether variations like those depicted in (9) are possible across languages.[16]

◁ internal organization of the syllable

(9) *Divergent syllable structures*




Let us follow the narrower, and therefore more illuminating path: accepting it as an axiom that such variation is not possible. All languages have a uniform syllable structure, potentially one of the above.

[14] More elaborate hierarchies are needed if we wish to account for phonotactic constraints, e.g., in English [pr] and [pl] exist syllable-initially, but [pn] does not. Assigning a sonority index to segments, or segment classes, would enable the analyst to formulate conditions on the necessary sonority distance between segments in syllables. But even these are not enough, a number of contraints (e.g., [pl] vs. *[tl], *[pw] vs. [tw]) cannot be explained in this way (cf. Selkirk 1984).

[15] Quite to the contrary: at the end of this chapter it will appear that all languages follow the simplest structure, CV.

[16] These three cases exhaust the possibilities unless association lines are allowed to cross.

In order to evaluate the proposals made about what this uniform structure should look like, we must first collect reasons for assigning internal structure to syllables. A hierarchical system of constituents is posited by linguists to show that some elements are more closely related than others. In the noun phrase *the black cat,* the words *black* and *cat* are within a constituent that does not contain *the* because their relationship is found to be tighter than that of either with the determiner. Also, in feature geometric representations of the internal structure of segments, those features or nodes are listed under a common node that pattern together, that is, spread together, or are deleted together. Syllable internal structure is thus expected to reflect that not all the segments within a syllable are peers of each other.

The prototypical syllable contains a consonant followed by a vowel (CV). Languages in which this type exhausts the syllable type inventory are not instructive of internal structuring of syllables at all. More complex systems allow closed syllables, CVC, and/or long vowels, CVV. In the case of such syllables, subdivisions make sense, in fact, they are obligatory if one accepts that constituents are maximally binary branching▷(Kayne 1984, Kaye et al. 1990). If so, the option depicted in (9a) is excluded on theoretical grounds. This 'flat' structure is employed by Kahn (1976) and Clements & Keyser — although the latter also posit a so called nucleus display, but their nucleus is not a daughter of the syllable (1983: 17). In itself it does not provide much information other than the location of syllable boundaries.[17] ◁ binary branching

The choice between the other two possibilities, (9b), in which the initial consonant enjoys more independence, and (9c), in which the initial consonant and the vowel are more closely related, is not

[17] Kahn chooses the nonlinear syllable tree of (9a) to provide for ambisyllabic consonants. With syllable boundaries inserted between segments ambisyllabicity would either be impossible — V\$C\$V indicates a syllable comprising a single consonant—, or, with labelled bracketing, unconstrained—besides $[_a \ldots V[_b C]_a V \ldots]_b$, the bracketing $[_a \ldots V[_b CC]_a V \ldots]_b$ would also be possible. The tree representation, together with the no-crossing-lines constraint allows the option of being ambisyllabic only for single consonants.

obvious.[18] Arguments can be based on two kinds of phenomena: syllable weight and phonotactic relations between adjacent segments. Vowel length in closed syllables, which may be related to both categories, will be discussed separately.

## 3.1 Syllable weight

Syllable weight depends on the number of segments in the second half of the syllable. The vowel always contributes to the weight of the syllable (i.e., it is always moraic). Long vowels, and in some languages syllables closed by a consonant, contribute two moras, thus make the syllable heavy, type (10a). There also exist languages where the weight of syllables depends exclusively on the vowel they contain, any consonant closing the syllable is ignored, type (10b). Let it be mentioned that there exist cases of a third system in which some closed syllables (those closed by a sonorant consonant) are heavy, others (closed by an obstruent) are light (Lithuanian and Kwakwala are well-known examples).[19]

◁ syllable weight

(10) *Two systems of syllable weight*

| | light | | heavy | example |
|---|---|---|---|---|
| a. | —V | —VC | —VV | English, Hungarian, Cairene Arabic |
| b. | —V | —VC | —VV | Lardil, Khalka Mongolian |

As opposed to syllable-final consonants, those preceding the vowel

[18] Nevertheless, the structure in (9b) is significantly more popular among theorists. McCarthy is one who suggests a tree like (9c) to explain the extrametricality of word-final consonants in Cairene Arabic (1979: 453).

[19] Ewen & van der Hulst (2001: 135) tentatively suggest that Dutch may exemplify a fourth type, in which closed syllables are heavy, but syllables with a long vowel are light. Kenstowicz claims that such languages do not exist (1994: 428). We will return to this issue in footnote 23.

never[20] influence the weight of the syllable. Interestingly, this asymmetry may be used to argue in both directions. One may say that the weightlessness of the syllable-initial consonant sets it off as an independent unit, whereas the fact that a syllable-final consonant is potentially weighty likens it to the vowel, which, recall, always carries weight. Accordingly, the internal structure of the syllable to be preferred is the one in (9b). The intermediate constituent containing the vowel and the consonant following it is called the rhyme. Adherents of this view will then say that only segments dominated by the rhyme may be moraic, those outside the rhyme — in the onset — never carry weight. However, this argument cannot hold for languages where closed syllables are not heavy, type (10b). So one either posits structure (9c) for these languages, at least for closed syllables (McCarthy 1979: 455), thereby giving up the axiom that syllable structure is uniform across languages, or the weight argument for the rhyme is lost. In other words, reference to syllable weight argues for the constituenthood of the rhyme only in those languages where closed syllables are heavy, type (10a), not in others where they are light, type (10b). ◁ rhyme

In fact, the selection of structure (9c) could also be justified by the same facts. Under this view weight is represented by the immediate constituents of the syllable. The weightlessness of the onset now follows from the fact that it always shares a constituent with the vowel, while the syllable-final consonant, the coda, is a separate unit, another mora (cf. Ewen & van der Hulst 2001: 129). Unfortunately, the parametric variation of languages with heavy vs. others with light closed syllables again cannot be explained in a straightforward way, since in the weighty-coda systems the immediate branching of the syllable node adds weight, in weightless-coda systems it does not. That is, weighty-coda languages can be formalized equally well with both structures (9b) and (9c), but both fail in the case of weightless-coda systems.

[20] If not never, then so marginally that the counterexamples (e.g., Everett & Everett 1984, Davis 1988) can, for the time being, be safely ignored as potential misanalyses (cf. Takahashi 1999, Goedemans 1996).

Thus, facts about syllable weight do not provide conclusive evidence either for the constituenthood of the onset and the nucleus with the exclusion of the coda, or for that of the nucleus and the coda with the exclusion of the onset.

### 3.2 Phonotactic constraints

◁ phonotactic constraints

Let us turn to phonotactic constraints▷then. The question again is whether there is a significant difference in the intimacy of the relationship of a vowel with the consonants on either of its sides. The right-branching structure in (9b) forecasts that the onset is independent of the rest of the syllable phonotactically, but the coda is not. Against such a view, Clements & Keyser state: 'cooccurrence restrictions holding between the nucleus and preceding elements of the syllable appear to be just as common as cooccurrence restrictions holding between the nucleus and following elements' (1983: 20). They then produce a wholly unconvincing list of English phonotactic constraints to support their cause. As the authors themselves admit, some items on the list mention minor constraints that could easily be taken as accidental gaps. 'Voiced fricatives [...] are excluded before /ʊ/': both voiced fricatives and /ʊ/ are among the less frequent segments of Modern English, that their combination is even rarer is not surprising. But the situation is worse: words like *bivouac, casual, usual, zucchini* do contain the allegedly impossible sequences.[21] Other constraints on Clements & Keyser's list are more extensive and do not involve exclusively the onset and the following rhyme. For example, the $\mathrm{sC_i\check{V}C_i}$ constraint refers to a coda [s] followed by an onset consonant, a vowel, and another consonant that may well be the onset of yet another syllable. That is, this is not a constraint on an onset and the following nucleus/rhyme. In fact, Törkenczy (1994) shows that this is a morpheme structure constraint and has nothing to do with syllables, with respect to this

[21] Northern England accents that have /ʊ/ where southern accents have /ʌ/ (e.g., *vulgar* [vʊlgə], *result* [rɪzʊlt], *thus* [ðʊs]) show clearly that this 'constraint' is the result of a historical process in the south that weeded out most occurrences of /ʊ/, i.e., the scarcity of such sequences is due to an independent sound change, not their ill-formedness.

constraint the status of $CC_iV\$C_i$ is the same as that of $CC_iVC_i\$$ (and as $C\$C_iVC_i$, for that matter). Yet other claims Clements & Keyser make are simply false ('in dialects distinguishing /aː/ from /ɒː/, the sequence /waː/ is excluded': British English is such a dialect with /wɑː/ in *Botswana, guano, qualm, quark*; '/Cl/ clusters are excluded before /ʊ/': *fluoride, influence,* etc.; '/Cr/ clusters are exluded before /er,or,ar/': *agrarian, arbitrary, contrary, librarian, prairie, sombrero, crore, registrar*).[22]

Kaye's view is the absolute opposite: he claims that any onset may combine with any rhyme, that is, there are no constraints whatsoever between the prevocalic part of the syllable and the rest (1985: 290; also cf. Pike & Pike 1947, Selkirk 1982). Even if some languages do exhibit onset–nucleus constraints (like *[wu], *[ji]), the robust pattern is that these are few and far between. This constitutes an argument against the left-branching structure in (9c). But it does not automatically justify the right-branching structure in (9b). If one is to argue for that, one has to show that the vowel and the consonant at the end of the syllable are related.

In fact, the interaction between a vowel and the following consonant is not as common as tacitly assumed by proponents of structure (9b). The coda constraints found in languages typically constrain the types of segment that can occur in coda position, irrespective of the preceding vowel, in fact, often depending on the following consonant.

Fudge notes that in English certain consonant clusters can only be preceded by short vowels, thus long vowels and diphthongs do not appear before noncoronal nasal+stop or stop+stop clusters (1969: 272f). Many other languages restrict the length of the vowel in a closed syllable, note however that this does not effect the quality

[22] In English, onset–nucleus constraints all involve consonant+glide onsets: any vowel can follow a glide, but Cj can only be followed by [uː] (and its r-influenced version), while Cw—where C is not [s]—cannot be followed by [uː] (or [aʊ] either, which incidentally is the continuation of Middle English [uː]).

of the vowel, but its quantity.[23] That is, it is not necessary to include the vowel and the following consonant in a single constituent to explain why no consonants may occur after a long vowel. We will return to this point in section 3.3.

◁ nucleus–coda interaction

Examples▷for genuine qualitative changes happening to a vowel because of the following consonant are breaking/broadening in English or nasalization in French. Consider the data in (11) taken from Giegerich (1992: 55, 63). They are somewhat simplified to illustrate the point. S stands for Scottish Standard English (SSE), A for General American (GA), and B for Standard Southern British English (SSBE).[24]

(11) *Plain and pre-R vowels in three accents of English*

a. plain

| | S | A | B |
|---|---|---|---|
| shoot | u | u | u |
| spoke | o | o | o |
| shot | ɔ | ɑ | ɒ |
| butt | ʌ | ʌ | ʌ |
| bid | ɪ | ɪ | ɪ |
| head | ɛ | ɛ | ɛ |

b. pre-R

| | S | A | B |
|---|---|---|---|
| sure | u | ʊ | ɔ |
| sport | o | ɔ | ɔ |
| short | ɔ | ɔ | ɔ |
| word | ʌ | ɜ | ɜ |
| bird | ɪ | ɜ | ɜ |
| heard | ɛ | ɜ | ɜ |

All three accents exhibit six different vowels before consonants other than [r], as shown in (11a). In pre-R position, however, only SSE has the same vowels, in GA we find three different vowels, in SSBE the contrast reduces to two only, (11b). This difference is accompanied by the fact that in SSE [r] is a full-fledged consonant, trill

[23] Claiming that it is tense vowels that are excluded here has a highly undesirable consequence: syllable weight would thus be dependent not on the quantity but the quality of the vowel. In fact, this is the solution to the weird pattern of Dutch: the stress placement algorithm sees closed syllables as heavy, long-vowelled ones as light, which is an absolutely unexpected pattern. If ‘long’ vowels are, in fact, tense and short, Dutch fits into the well-established categories described in (10) (Ewen & van der Hulst 2001: 134f).

[24] Formerly known as Received Pronunciation (RP).

or flap or fricative (Wells 1982: 411), while in SSBE it merges with the preceding vowel and disappears altogether. In General American nonprevocalic R is realized as an r-coloured vocoid (Wells 1982: 490). Thus, R influences the preceding vowel only if it vocalized, that is, when it becomes part of the nucleus, but not when it is consonantal, that is, when it is in the coda. (Also cf. Harris 1994: 257ff.) French front vowels are lowered by a following nasal, but again only when the nasal vocalizes and merges with the preceding vowel (cf. *fin* [fɛ̃]∼*fine* [fin] 'fine masc.∼fem.', *un* [œ̃]∼*une* [yn] 'one masc.∼fem.').

Without an exhaustive survey of vowel–consonant sequences across languages, we may risk the claim that phonotactic constraints are no more frequent between nucleus and coda than between onset and nucleus (Harris 1994: 64). I agree with Clements & Keyser, after all: there are no significant phonotactic constraints between consonants and vowels. Such constraints hold between members of diphthongs, long vowels, as well as between members of consonant clusters, irrespective of whether they are in the same syllable, or not.

### 3.3 Closed syllable shortening

Another apparent argument for structure (9b), the onset vs. rhyme split of the syllable, is the phenomenon of closed syllable shortening.▹ In a number of languages, a syllable is either closed or has a long vowel, but not both. Italian non-final stressed syllables exemplify the phenomenon,[25] some data are given in (12).

◃ closed syllable shortening

[25] The Italian data may also be analysed as open syllable lengthening: the vowel of an open syllable is lengthened so that stressed syllables are always heavy. Crucially, the vowel of a closed syllable is not lengthened.

(12) *Closed syllable shortening: Italian stressed syllables*

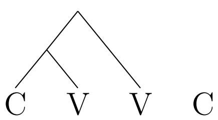
a. open syllable → long vowel
fā.ce 'torch', fā.ma 'fame', fā.me 'hunger', fā.ro 'lighthouse', fā.se 'phase', fā.ta 'fairy', fā.va 'bean', fī.bra 'fiber'

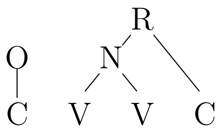
b. closed syllable → short vowel
făb.bro 'smith', făg.gio 'beech', făl.ce 'sickle', făl.co 'falcon', făn.go 'mud', făn.te 'pedestrian', făs.to 'luxury', fĭr.ma 'signature'

The syllable boundary is marked in the words to show that when it follows the vowel immediately, the vowel is long, when it does not, the vowel is short. This distribution can be explained by claiming that the rhyme constituent must contain maximally (in the case of shortening in closed syllables) or exactly (in the Italian case discussed here) two segments.

However, structure (9c) explains such a restriction just as well: this tree contains a single position following the onset and the first part of the nucleus, thus there is no place for a coda consonant after a long vowel in this tree, as (13a) shows.

(13) *Left- and right-branching syllable trees and closed syllable shortening*

a. 


b. 


In fact, the onset–rhyme model cannot exclude a closed syllable containing a long vowel by reference to binary branching alone: as (13b) shows, a rhyme can contain three segments *and* observe binary branching (cf. Harris 1994: 163f). To exclude this structure▷— ◁ binary branching
and explain closed syllable shortening — Charette (1989: 161ff) borrows the notion of c-command from syntax, while Kaye et al. (1990:

199) claim that the three-member rhyme constituent of (13b) is impossible because the head cannot properly license all its dependents.[26]

Structure (9c) — as (13a) shows — falls short of explaining the phonotactic constraints evidently holding within a complex nucleus: diphthongs in a language are typically a small subset of all the possible combinations of short vowels. But then, as we have seen, phonotactic constraints hold even between segments that do not belong to the same syllable: consonant clusters flanking a syllable boundary often are so related.

To conclude, we have not found compelling evidence for either the onset vs. rhyme split or the onset-cum-nucleus vs. coda split within the syllable. The theoretician now has two options: either he reverts to the flat syllable tree of (9a), which undesirably violates binarity, or he asks whether the syllable node exists at all. What phonotactic constraints call for is primarily vowel–vowel and consonant–consonant communication, which is clearly independent of syllables. By dispensing with the syllable as a constituent in the prosodic hierarchy, we do not automatically make syllable-related explanations of phonological phenomena impossible. In what follows, I will discuss an alternative model of syllable structure that lacks any hierarchical structure, but is still capable of expressing most of what standard syllable trees can.

## 4 Vowelless syllables

It has already been hinted at that word edges exhibit peculiar behaviour with respect to syllable structure. For example, two obstruents (excluding the weirdly behaving [s]) cannot be adjacent within a syllable except when they are at the edge of a word. The definition of the universal syllable in (4) predicts that such a cluster will be separated by a syllable boundary and this does not depend

[26] Note that the trouble the onset–rhyme model faces stems from labelling both the nucleus and the rhyme. The original tree in (9c), without these labels, is just as restricted. In it the rhyme can contain two segments: either two vowels, or a vowel and a consonant.

on the position of the cluster within the word. That is, words like the following must contain a syllable boundary at the dot: *pas.t, sof.t, fac.t, s.top,* etc. Note that now two obstruents are not adjacent within a syllable even at the edge of a word. Recall that in some languages after such a syllabification epenthesis ensues (e.g., in Yiddish [g.dɔ.rɪm]>[gɪ.dɔ.rɪm]). But in the above English examples epenthesis does not occur.

In the case of another prosodic unit, the foot, it is accepted that word boundaries may not coincide with foot boundaries. A foot may span across a word boundary (e.g., the first foot in |*Sám's a*|*lóne*|), and it may also occur that we only find the second half of a foot at the beginning of a word, e.g., in *a*|*lóne,* where the second syllable initiates a foot, while the first, unstressed syllable lacks a head. Such a foot is called a degenerate foot.

Since syllables are also prosodic units, and thus their segmentation should be largely independent of semantic units (i.e., words), I accept Aoun's (1979) proposal, adopted and elaborated by Kaye (1990a), that such vowelless (i.e., headless) syllables are degenerate syllables.[27] Thus the last consonant of *pas.t* forms a degenerate syllable that lacks a vowel. Claims like 'there are no well-formed syllables in any language that lack an overt nuclear segment on the surface' (Hayes 1989: 286), or 'in all languages, syllable edges correspond with word/utterance edges' (Blevins 1995: 209) have no theoretical status, they describe beliefs. One has to support such a claim with arguments. The onset of a syllable may be empty in a large set of languages, it is not clear why we should accept without

◁ vowelless syllable

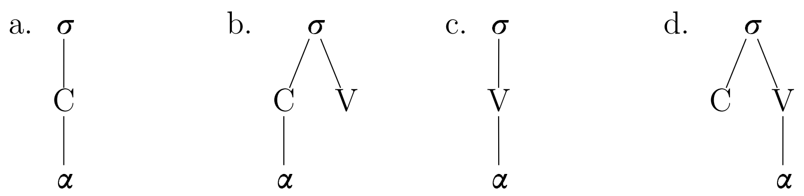
◁ degenerate syllable

[27] Aoun restricts the use of degenerate syllables to the last, 'extra' consonant of superheavy (-VVC or -VCC) syllables, while Kaye extends the notion to any word-final consonant. The fact that a word-final consonant does not influence the weight of the last syllable in many languages, English among them, was formalized by labelling these consonants extrametrical (Hayes 1982). The phonotactics of word-final consonants also shows their independence (Itô 1986, Myers 1987, Borowsky 1989). Harris argues extensively that the set of English word-final consonant clusters is remarkably similar to that of intervocalic, *heterosyllabic* clusters (1994: 66ff), Harris & Gussmann (1998) show that the western tradition of syllabifying word-final consonants as a coda is probably wrong.

argument that the same status is denied of nuclei. One could argue that the nucleus is the head of the syllable, but headless feet are accepted and headless structures occur abundantly in syntax, too. If the sentence is analysed as a complementizer phrase, the head complementizer is very often missing from the surface. Another possible argument could be that a syllable is unpronounceable without a nucleus. But in words, nucleusless syllables always cooccur with others that have a nucleus: words always contain some syllabic segment, it is clitics that do not and thus need the phonological support of an adjacent word. In fact, as we will see, the number of consequent nucleusless syllables is strictly constrained in languages. The general aversion towards vowelless syllables stems from the fact that the view of the syllable as a wave of sonority seems impossible without the peak of sonority that the vowel represents. The definition of the universal syllable in (4) shows that this is not the case: the empty vowel between two obstruents can be detected, just like the empty consonant between two vowels that create hiatus, as both constitute segment pairs of equal sonority.

There are two senses of a syllable being vowelless (or consonantless), as shown in (14), where $\alpha$ stands for any segment.

(14) *Emptiness in a syllable*




Both (14a) and (14b) represent a degenerate syllable, but while the first syllable contains a sole consonant (this is Aoun's conception), the second is a prototypical syllable, it is just that its nucleus is not associated with any melodic material: it is empty. The same is true of the two potential representations of the onsetless syllable in (14c) and (14d). In what follows, I will adopt a version of representations

(14b) and (14d).[28]

As mentioned in footnote 27, the independence of word-final consonants could be solved without the introduction of degenerate syllables, by labelling them extrametrical. Myers says: 'An extrametrical element is inaccessable to the phonology, i.e., it can neither condition nor undergo phonological operations' (1987: 487). In a footnote Myers adds that the inaccessability refers only to the extrametrical element, not to its features. This is necessary to explain the phonotactic constraints holding between two word-final consonants (cf. Harris 1994: 66ff). Still, if a word-final consonant is extrametrical then we do not expect to find -V̆C# sequences in English, since word-final stressed short vowels do not occur. This paradox does not arise with the degenerate-syllable analysis.

In what follows, we will extend the distribution of degenerate syllables in an extreme way: any consonant will be seen as the 'onset' of a 'syllable' either with or without a vowel.

## 5 Syllables without constituents

Let us begin with quoting[29] the French academician Charles Pinot Duclos (1704–1772):

> Pour distinguer la syllabe réelle ou physique, de la syllabe d'usage, il faut observer que toutes les fois que plusieurs consonnes de suite se font sentir dans un mot, il y a autant de syllabes réelles qu'il y a de ces consonnes qui se font entendre, quoiqu'il n'y ait point de voyelle écrit à la suite de chaque consonne: la prononciation suppléant alors un *e* muet, la syllabe devient réelle pour l'oreille, au lieu que les syllabes d'usage ne se comptent que par le nombre des voyelles qui se font entendre et qui s'écrivent. [...] Par exemple, le mot *armateur* seroit en vers de trois syllabes d'usage, quoiqu'il soit de cinq syllabes réelles, parce qu'il faut suppléer un *e* muet après chaque *r*; on entend nécessairement *arematteure.* *Bal* est monosyllabe d'usage, et dissyllabe physique. (1803: 407f)

[28] The syllable in (14d) is not considered degenerate, just like a monosyllabic foot is not a degenerate foot. Both are, nevertheless, marked structures compared to a CV syllable or a strong–weak foot.

[29] Thanks to András Cser for calling my attention to this comment.

Whenever it is felt that consonants are adjacent in a word, each consonant constitutes a syllable (réelle), even if there is no vowel following it. The traditional sense of the syllable (la syllabe d'usage) counts only pronounced vowels. Duclos claims that any consonant that is not followed by a vowel constitutes a degenerate syllable: the superficially trisyllabic word *armateur* is in fact to be syllabified as five syllables *a.r.ma.teu.r,* shown in (15) in current nonlinear terms.

(15) *The representation of* armateur

$$\begin{array}{cccccccccc} \mathrm{C} & \mathrm{V} & \mathrm{C} & \mathrm{V} & \mathrm{C} & \mathrm{V} & \mathrm{C} & \mathrm{V} & \mathrm{C} & \mathrm{V} \\ & | & | & & | & | & | & | & | & \\ & \mathrm{a} & \mathrm{r} & & \mathrm{m} & \mathrm{a} & \mathrm{t} & \mathrm{\oe} & \mathrm{r} & \end{array}$$

As for vowel clusters, Duclos says:

> A l'égard de la diphtongue, c'est une syllabe d'usage formée de deux voyelles, dont chacune fait une syllabe réelle, *Dieu, cieux, foi, oui, lui.* (1803: 408)

So, just like consonant clusters, diphthongs also form two underlying 'syllables', that is, two vowels separated by an empty consonant.[30] Almost two and a half centuries later,[31] Lowenstamm (1996) arrives at the same conclusion: the phonological skeleton consists of strictly alternating consonantal and vocalic positions, as shown in (15). Note that if the simplest sonority hierarchy (C→V) is applied to the universal syllable in (4), then exactly this 'syllabification' follows. As regards their position on the skeleton, two vowels or two consonants are never adjacent. The model also lacks any branching above the skeleton: syllabic constituents are missing. This conception of prosodic organization solves a number of issues left unresolved above: whether the onset or the coda belongs with the nucleus, or whether syllabicity is lexically marked.

[30] Perhaps Duclos is not right about the representation of French light (monomoraic) diphthongs, his idea is still very avant-garde.

[31] Although I here quote from the 1803 edition, Duclos's comments on the Port Royal grammar were first published in 1754.

However, several other questions arise in connection with this extremely simplified syllable structure. Some of them are listed in (16), and discussed in the following sections.

(16) *Questions about the strict CV view of the skeleton*

a. how are empty V positions controlled?
b. why is the length of consonant clusters limited in natural languages?
c. why is the length of vowels limited in natural languages?
d. why are some consonant clusters possible, and others not?
e. how are long vowels/diphthongs and hiatus distinguished?

## 5.1 Consonant clusters

If consonant clusters▷do not form constituents whose size can be limited, there has to be some alternative method of constraining them. In a syllable-tree model, the size of a syllabic constituent containing consonants may be controlled by the number of branches that it may have. In the strict CV approach the same effect is achieved by controlling the number of subsequent empty vowels. Emptiness is marked:[32] the occurrence of an empty vocalic position must be sanctioned. Kaye et al. claim that it is the pronounced vowel of the next syllable that sanctions the nonpronunciation of an empty vowel▷ by a relation they call (proper) government▷(1990: 219). The pairs in (17) taken from various languages show that the appearance of the vowel▷indeed depends on the appearance of the following vowel. (Sources of the data: Brockhaus 1995: 214 for German, Gussmann 2007: 185 for Polish, Kaye 1990b: 140 for Moroccan Arabic, Kenstowicz 1994: 129 for Somali, my own competence for the other languages.)

◁ consonant clusters
◁ V-to-V government
◁ proper government
◁ vowel–zero alternation

[32] Kaye & Lowenstamm (1981) and Cairns & Feinstein (1982) discuss markedness and syllable structure, Szigetvári (2006) discusses the relation of markedness and complexity.

(17) *Vowel–zero alternations across languages*

a. English: *simple* [sɪmpəl] ∼ *simpler* [sɪmp•lə]
b. French: *j'amène* [ʒamɛn] 'I bring' ∼ *amener* [am•ne] 'to bring'
c. Latin: *pater* 'father' ∼ *pat•ris* 'gen.'
d. German: *Eben* 'level' ∼ *Eb•nung* 'levelling'
e. Polish: *łeb* 'head' ∼ *ł•ba* 'gen.'
f. Hungarian: *majom* 'monkey' ∼ *maj•ma* 'his/her monkey'
g. M Arabic: [tan k•dɨb] 'I lie' ∼ [tan kɨd•buː] 'we lie'
h. Somali: [nirig] 'baby female camel' ∼ [nir•go] 'pl.'

In each morphologically related pair in (17) the vowel in __C# position is pronounced, but when a suffix provides a pronounced vowel after it, that is, it comes to be in __CV position, it remains silent (marked by a bullet). English also has a large degree of optionality in its vowel–zero alternations: *stationary* may be pronounced [steɪʃənəri], [steɪʃ•nəri], or [steɪʃən•ri], but not *[steɪʃ•n•ri]. We see that two successive silent vowels are not allowed (also cf. *[tan k•d•buː] in (17g)), the first of them would not be sanctioned because the second, being unpronounced cannot govern it: only pronounced vowels can govern.

Two conclusions follow: (i) there are no consonant clusters word finally (since there is no following pronounced vowel to govern the one in the cluster) and (ii) there are no three-member consonant clusters intervocalically (since two successive silent vowels are not allowed). While true of some languages, others falsify these conclusions. Clearly, V-to-V government cannot be the only source of consonant clusters in languages.

Kaye notes that in Moroccan Arabic members of a class of nouns fit one of two templates: CVCC or CCVC, exemplified by [mɨrd] 'sickness' and [sdɨr] 'chest' (1990b: 149). The choice of the template depends on the last two consonants: if they exhibit the right 'sonority' relation[33] — falling sonority —, the two consonants are super-

[33] Whether it is sonority or some other property of sounds—charm (Kaye et al. 1985, 1990) or complexity (Harris 1990) — that controls their sequencing is a question I do not address here.

ficially adjacent, if not, they are separated by a vowel. This difference is due to the fact that if consonant A is capable of governing consonant B, then B cannot govern A (see Kaye et al. 1990 for details). In the specific example, [d] can govern [r], but not vice versa. This C-to-C government▷ relation creates a governing domain within ◁ C-to-C government
which the vowel remains mute ($V_2$ in (18a); cf. Kaye 1990a: 322). The pronunciation of the first vowel is a function of the pronunciation of the second: when $V_2$ is mute $V_1$ is pronounced, when $V_2$ is pronounced — because there is no C-to-C governing domain above it — $V_1$ is governed by it, hence it remains mute, (18b). Government is indicated by a blunt arrow.[34]

(18) *The two nominal templates of Moroccan Arabic*

a.

| C | $V_1$ | C | $V_2$ | C |
|---|---|---|---|---|
| \| | | \| | | \| |
| m | | r | | d |

b.

| C | $V_1$ | C | $V_2$ | C |
|---|---|---|---|---|
| \| | | \| | | \| |
| s | | d | | r |

Thus, the two nominal templates of Moroccan Arabic are in fact a single template. The difference in the superficial order of Cs and Vs is an effect of the governing relations between the segments that are associated to the template.[35] Note that government operates from right to left: it is [rd] that creates a C-to-C governing domain, not [dr].[36] The direction of government was already tacitly established for V-to-V government: vowel syncope depends on the presence of a following, not on that of a preceding vowel. Also note that this effect of consonant clusters is not observed word initially (we

[34] C-to-C government is drawn below the intervening V, because this type of government is only possible across an empty V: the arrow graphically inhibits association to melody.

[35] Amimi (1996) shows that the distribution is more complicated with competing alternants. He claims that the basic CCVC template—which lacks C-to-C government—may surface as CVCC, if C-to-C government is possible.

[36] These facts also imply that the governing relation between skeletal positions is not part of the template, but depends on the melodic content associated with the given position.

find forms like [dreʕ] ‘se propager’ vs. [rdeʕ] ‘mettre quelqu’un à sa place’; Abdellah Chekayri, p.c.).[37]

It is also to be observed that no vocalic melody is associated to the template in (18). Recall the crudest sonority hierarchy mentioned in section 2.3: C→V. Dienes & Szigetvári (1999) claim that this hierarchy represents the top and bottom of the ideal sonority wave in speech: loudness (V) and silence (C), it is the alternation of these two that the skeleton encodes. The prototypical consonants, plosives, are a brief period of silence, the prototypical vowel is a peak of loudness. That is, a V slot is loud, and a C slot is silent, without any melody attached to it. The effect of government▷can now be generalized as in (19).

◁ effect of government

(19) *The effect of government*

Government spoils the inherent property of its target.

The loudness of a governed V position is spoilt: such a vowel remains silent. The ability of a vowel to govern and license is also extinguished by government. For a C position government tends to spoil its silence: the first consonant in a C-to-C governing domain is typically a sonorant. Languages may be more tolerant in the types of consonant cluster they allow — besides the unmarked sonorant+obstruent cluster others may occur in some languages —, but then C-to-C government is not the only configuration for surface consonant clusters, as (18b) shows. Thus, word finally we expect only consonant clusters that are created by C-to-C government▷(since there is no V position to govern the V enclosed within the cluster). We also have a means of modelling a word-internal three-member

◁ C-to-C government

[37] The notion of C-to-C interaction has alternative roles in later analyses. Cyran & Gussmann (1999), for example, call it interonset government, but they apply it in the reverse direction. Scheer (1996, 2004) applies right-to-left infrasegmental licensing. These models differ from Kaye’s implementation in that they silence the vowel within a rising-sonority cluster, i.e., a branching onset. C-to-C government as defined here is thus a return to Kaye’s original version.

consonant cluster now,[38] as shown in (20): this is the only way two successive empty V positions can be silenced, since the first empty V position cannot be governed (silenced) by the second. (A further option is excluded in (24).)

(20) *Word-internal three-member consonant cluster*

C V C V C V
n t l ɑ

Thus, consonant clusters are of two types: those that are created by government between two C or between two V positions. The differences between the two types are discussed in detail elsewhere (Szigetvári 2007).

Now that any C is followed by a V, we seem to lose the original generalization about syllable-initial and syllable-final consonants, that is, onsets and codas. In fact, the two environments can still be captured: an onset is a C position followed by a pronounced V position, a coda is one that is not followed by a pronounced V position.

## 5.2 Vowel clusters

If the representation of long vowels▷ and diphthongs involves an empty C position between the two V parts, these and hiatus▷ become identical. In a hierarchical framework,▷ they are distinguished by a branching structure, (21a) vs. (21b), or also an empty C slot, an onset, (21a) vs. (21c). (Since they may here branch, I use the labels O and N, not C and V, for the consonantal and the vocalic parts of the representation.)

◁ long vowel
◁ hiatus
◁ diphthong

(21) *Diphthong vs. hiatus (traditional view)*

a. N — V V
b. N N — V V
c. N O N — V V

[38] Four-member clusters, as in English *i*[nstr]*ument, e*[kspl]*ain,* cannot be dealt with. These all involve [s], the *enfant terrible* of phonotactics (cf. Kaye 1996).

In the strict CV model, neither of the first two options are available, since there are no branching structures, and any two V positions are separated by a C position. Long vowels/diphthongs and hiatus cannot thus be distinguished by their skeletons, some further mechanism is needed.

Prosodic licensing▷is a widely accepted notion in phonology (Itô 1986, Goldsmith 1990, Harris 1997). It is meant to be an explanation of why onsets allow more contrasts than codas, or stressed nuclei than unstressed ones: the former are better licensed than the latter. It is also consensual that vowels license the preceding consonant (e.g., Kaye et al. 1990, Charette 1992, Harris 1997, Steriade 1999, Cyran 2010). The strict CV model applies licensing, besides government, as a relation between positions on the skeleton: any V position that is pronounced licenses the C position before it (Ségéral & Scheer 2001). Following the standard assumption, the effect of licensing▷is given in (22).

◃ prosodic licensing

◃ effect of licensing

(22) *The effect of licensing*

Licensing supports the melodic content of the licensed position.

However, contrary to standard assumptions, licensing — as Ségéral & Scheer conceive of it — is not a *sine qua non* for all portions of the representation: C positions that are not followed by a pronounced vowel are not licensed. This results in their reduced capacity to maintain melodic elements (i.e., contrasts), but does not necessarily inhibit expression of the given position. That is, licensing as conceived of here is not prosodic licensing (sanctioning the presence of a position) or government licensing (sanctioning a consonant cluster), but exclusively autosegmental licensing▷(sanctioning melodic elements associated with a skeletal position).

◃ autosegmental licensing

It is also important to see that a licensor is not necessarily more prominent than its licensee — unlike, for example, Harris (1990, 1997) sees it —, it is merely a pronounced vocalic position. Dienes & Szigetvári (1999) contend that two pronounced V positions surrounding an empty C position may be lexically specified to be in a licensing relationship.[39] This V-to-V licensing▹domain defines long vowels and diphthongs, as shown in (23). ◃ V-to-V licensing

(23) *Diphthong vs. hiatus (strict CV view)*

a. V C V
   | |
   ɑ u

b. V C⇐V
   | |
   ɑ u

Like C-to-C government, V-to-V licensing is also possible only across an empty position. But since a C position is silent unless it is forced to be pronounced (by government or by being associated with melodic elements), there is no need to sanction empty C positions. Thus complex nuclei and hiatus are only different in the presence and absence of licensing between two neighbouring V positions.

The standard explanation for the nonexistence (or marginality) of three-way vowel length contrast is the absence of ternary branching▹syllabic constituents, nucleus in our case. The framework introduced here seems to allow such structures, as shown in (24). ◃ ternary branching

(24) *Ternary clusters*

a. V C V C V

b. C V C V C

To encode the observation that such vowel and consonant strings do not occur in natural language, the possibilities above must be excluded. As for (24b), one could claim that being governed deprives a position from its governing capacity — as was claimed for V positions above —, but licensing certainly does not do so. If it did, a consonant before a long vowel would be unlicensed, which is an absurd possibility: it would entail that prevocalic consonants differ in their readiness to lenition depending on the length of the vowel. Instead, Dienes & Szigetvári (1999) call C-to-C government and

[39] Yoshida (1993) and Ritter (1995) analyse the long vowels of Palestinian Arabic and Hungarian, respectively, as constituting a left-to-right governing domain enclosing an empty onset.

V-to-V licensing burial domains (burying the enclosed empty V or C position), and stipulate a constraint inhibiting burial domains from overlapping. This shows that abandoning constituency prevents the analyst from using well-established notions like binary branching in syllabic constituents. On the other hand, it is a distinct advantage of the treeless model that it creates relations between segments that could not be related in syllable trees. Recall, for example, that phonotactic contraints link a coda and the following onset. This sequence cannot be made into a syllabic constituent in any traditional hierarchical framework. In the present treeless model, however, this relation is not different in kind of that holding a long vowel or diphthong together. Although Kaye et al.'s (1990) theory posits very similar relations between skeletal positions as here, it also contains syllabic constituent trees. Since the relations and the constituents encode the same facts, Takahashi (1993) rightly points out that one of the two devices is superfluous.

## 5.3 Intervocalic consonants

Recall, V-to-V government may be established between a pronounced and an empty V position, i.e., one which is not associated with any melody. The nonalternating vowels in (17) are lexically associated. The following vowel cannot govern a vowel associated with melody (Ségéral & Scheer 2001). In this case, it governs the consonant between the two vowels, as shown in (25).▷ ◁ V-to-C government

(25) *Intervocalic consonants are governed*

$$\begin{array}{ccc} \mathrm{V} & \mathrm{C} & \mathrm{V} \\ | & | & | \\ \text{ɑ} & \mathrm{t} & \text{ɑ} \end{array}$$

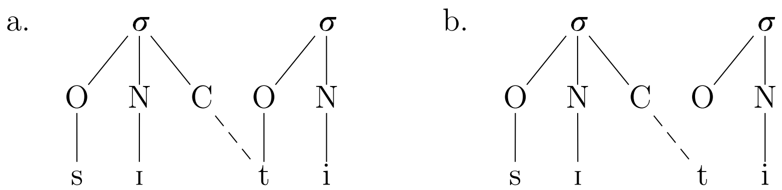
Since government spoils the inherent properties of its target (cf. (19)), the muteness of the C position in our case, the prediction is that intervocalic consonants will lose their muteness and become more vowel-like, more sonorous. This status of intervocalic consonants enables us to evade debatable notions of syllable theory: coda capture,▷ which may be implemented either as ambisyllabicity or as complete resyllabification.

◃ coda capture, ambisyllabicity, resyllabification

It is a well-documented fact that consonants in the coda of a syllable are 'weaker' than those in the onset: the latter are (better) licensed by the vowel that follows, the former are not (so well licensed), since no vowel follows them. This is manifested in two interrelated phenomena. Languages may limit the set of consonants — or rather the set of phonological features — that is allowed to appear in codas (by so-called coda constraints,▷ cf. Itô 1986, Borowsky 1986), but similar restrictions are very unusual for onsets. Coda consonants are also more prone to undergo lenition than onset consonants. Intervocalic consonants, however, are often also subject to lenition: flapping in several accents of English is an example. To synchronize lenition sites and syllabic affiliation several authors have proposed that core syllabification (where an intervocalic consonant is syllabified to the right as an onset) is followed by a process that assigns a (post-tonic) intervocalic consonant to the coda (or alternatively to both the coda and its original onset), that is, a (stressed) syllable captures the onset of the following syllable to make it its own coda (e.g., Kahn 1976, Selkirk 1982, Gussenhoven 1986, Rubach 1996). The change is illustrated by the ambisyllabicity, (26a), and the resyllabification, (26b), of the intervocalic [t] in *city.*

◃ coda constraints

(26) *Ambisyllabicity and resyllabification in* city




The idea of coda capture is theoretically uneasy. Ambisyllabicity — introduced by Hockett (1955: 52) as interlude — itself is highly problematic (cf. Selkirk 1982, van der Hulst 1985, Harris 1994, Jensen 2000), challenging the idea of proper bracketing.▷ The total resyllabification of an onset consonant into the coda of the preceding onset

◃ proper bracketing

works against onset maximalization, an apparently universal principle. Furthermore, the reasoning is circular: intervocalic consonants are resyllabified only to explain why they undergo lenition.[40]

A proposal to solve the problem of ambisyllabicity is to analyse such consonants as virtual geminates, a consonant that is represented as it were a geminate, but pronounced short. For example, van der Hulst (1985) argues that an intervocalic consonant following a short/lax vowel in Dutch is representationally a geminate, hence all such vowels are in a closed syllable. Now the consonant belongs to two skeletal positions, which can belong to two syllables respecting proper bracketing. The lenition facts of English, however, do not allow such an anaysis. There is no difference in flapping between the two types of [t], that following a long/tense and a short/lax vowel: *treaty* and *city* exhibit the same kind of flapping. In addition, if the latter had a virtual geminate, geminate integrity would make lenition outright unlikely. So the environment posited to condition lenition would be one where lenition were least likely.

◁ virtual geminates

◁ geminate integrity

To unify the two types of lenition environment, Harris (1997) advocates an intricate system of lenition paths starting out from a stressed nucleus, which licenses the unstressed nucleus in its foot, both nuclei licensing their onsets, and onsets licensing the preceding coda. The further down the licensing path, the weaker a position: this renders the onset of an unstressed syllable weaker than that of a stressed syllable. That is, the weakness of a coda compared to an onset is a consequence of licensing relations, not of syllabic affiliation. There is thus no need to resyllabify anything. However, Harris's model predicts (i) that the onset of *any* unstressed syllable will be weak and (ii) that intervocalic lenition is restricted to nonpretonic onsets. Neither prediction is right: while the [t] in *city* is flapped, that of *filter, mister, actor* is not. In fact, it is only

[40] The argument that coda capture is also necessary because words, and hence syllables, do not end in a short vowel in English is a weak one: if this were true, a word like *ferry* could not be syllabified at all. Neither [fɛ], nor [fɛr] is a possible word-final sequence in English, if the latter were syllable final, it would be a unique case where [r] occurred in the coda.

intervocalic[41] [t] and [d] that is flapped in English, exactly as (25) predicts. Furthermore, although English flapping happens only before an unstressed vowel, in other systems, for example, in Tuscan Italian, intervocalic stops spirantize irrespective of stress (e.g., *abete* Standard Italian [aˈbeːte], Tuscan Italian [aˈβeːθe]; Marotta 2008: 242). Such a language cannot be analysed in a framework that blames lenition on the distance of the licensing stressed vowel. The government-based lenition of consonants shown in (25), on the other hand, predicts sonorization of intervocalic consonants irrespective of stress, catering for the Tuscan pattern, but not for the English one. Ségéral & Scheer (2008: 507ff) and Szigetvári (2008: 581) propose two different means of accounting for the absence of lenition pretonically. Although this way an extra mechanism is involed in the analysis, the same mechanism is needed anyway to explain why syncope is also impossible before a stressed vowel (*memory* [ˈmɛm(ə)ri] vs. *memorize* [ˈmɛm*(ə)ˌraɪz]),[42] since — as is claimed here — both the lenition of an intervocalic consonant and syncope are the effect of the government coming from the following vowel.

## 5.4 Syllable weight

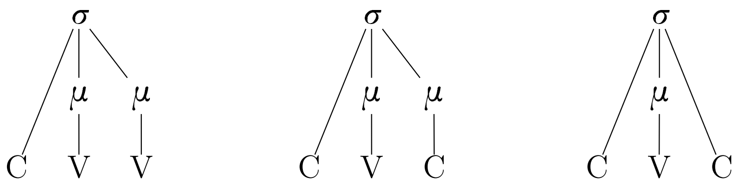
As we have seen in section 3.1, the internal constituency of syllables cannot adequately describe the two basic weight systems of (10).▷Replacing constituents like nucleus and coda with moras (Hock 1986, Hayes 1989) was an answer to this shortcoming. Theories applying moras distinguish the two weight systems of (10) by assigning a mora to both the vowel and the consonant following it, (27b), or only to the vowel, (27c). ◁ syllable weight

[41] The [t] in *party* and *twenty* is intervocalic following a rhotic and a nasalized vowel, respectively.

[42] The absence of pretonic syncope cannot be explained by stress clash avoidance, since syncope does not occur even if it would not result in stress clash: e.g., *hùllab*(a)lóo, mèthod*(o)lógical*.

(27) *Weight indicated by moras*

a. long V heavy =(10a, b)   b. coda heavy =(10a)   c. coda light =(10b)




A syllable containing a long vowel is always heavy, since vowels are invariably assigned a mora, (27a). The variability is in whether the coda is or is not assigned one. Scheer & Szigetvári (2005) claim that it is arbitrary that a mora may be assigned to a coda consonant, but not to one in the onset. This fact, however, follows from the strict CV model, where exclusively V positions are moraic. The difference between (10a) and (10b) depends on whether empty V positions are counted or not. Since what is traditionally called an onset consonant is always followed by a pronounced vowel it is always in a CV pair that is moraic, its weight being attributed to the vowel. Codas, on the other hand, are followed by an empty vowel. If this empty V position is counted (in weighty-coda languages), the coda consonant appears to carry a mora. Actually C positions are never moraic, when they appear to be, the illusion arises because the following empty V position participates in the mora count. This solution is less arbitrary because the same type of objects, V positions, contribute to syllable weight. It is not conceivable that in some language prevocalic consonants are moraic, but others are not. This nonattested scenario is not excluded by the moraic model. (In (28) unpronounced positions are symbolized by a lowercase letter, moraic V positions are underlined; the three representations parallel those of (27).)

(28) *Weight indicated by CVs*

| a. long vowel =(10a), (10b) | b. empty V counts =(10a) | c. empty V ignored =(10b) |
|---|---|---|
| C $\underline{\mathrm{V}}$ c $\underline{\mathrm{V}}$ | C $\underline{\mathrm{V}}$ C $\underline{\mathrm{v}}$ | C $\underline{\mathrm{V}}$ C v |

## 6 Conclusion

This chapter argues that (i) syllable structure is a useful notion in the description and explanation of phonological phenomena, but (ii) the representation of syllable structure by constituents visualized in trees runs into difficulties. Instead, a severely reduced model was offered applying a skeleton of strictly alternating consonantal and vocalic positions, representing muteness and loudness, respectively. The relations between segments are modelled by two forces, licensing and government. It is shown that this framework establishes relations only between segments that do exhibit phonotactic constraints, irrespective of whether they are tauto- or heterosyllabic. We get a noncircular explanation for why intervocalic consonants may undergo lenition. The strict CV model also offers an explanation for why onsets do not contribute to the weight of the syllable, while codas potentially do, instead of simply encoding this fact in the representation.