Syllables and Syllabification

David R. Mortensen

April 2, 2024

In almost all languages, sounds group, syntagmatically, into clusters that we call SYLLABLES. Syllables are sequences with:

- A SONORITY[1] peak
- Level or rising sonority up to that peak
- Level or falling sonority after that peak

The structure of a syllable is show in Figure 1. The structure shown in

[1] Definitions here are **circular**. Sonority is the degree to which a sound is a good NUCLEUS for a syllable. This correlates with the loudness of the sound. Low vowels are the most sonorous sounds, followed by mid and high vowels, then glides, then approximants, then nasals, then fricatives, then—finally—plosives.

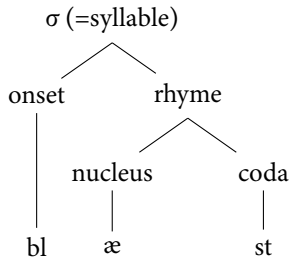


Figure 1: The structure of the syllable /blæst/ (*blast*).

Figure 1 was developed by Chinese linguists of the Sui Dynasty (around 600 CE) and the subsequent Tang Dynasty (as exemplified in the rhyme books called the *Qieyun* and *Guangyun*). This structure was later rediscovered by American linguists like Charles Hockett. There are many other theories of the syllable, but this is the most widely accepted one.

Hockett ties together many of the themes of this course. He was the one who divided morphological theories into Item-and-Arrangement, Item-and-Process, and Word-and-Paradigm. He proposed the first probabilistic model of morphology using a form of Hidden Markov Models, and collaborated briefly with Claude Shannon before falling ill.

Notation and Examples

In phonology, we often show syllable boundaries with period/full-stop:

(1) a. lin.guís.tiks /liŋ.ˈgwɪs.tɪks/
    b. in.des.crí.ba.ble /ɪn.dəs.ˈkraj.bə.bl̩/

Note that nuclei are usually vowels (sometimes MONOPHONTHONGS—single vowels—and sometimes DIPHTHONGS—a vowel plus a glide). However, they may also be consonants as in *indescribabl**e***.

Evidence for Syllables

Syllables define important environments for phonological patterns.

(2) a. Stress is assigned to syllables
    b. METRICAL FEET are defined in terms of syllables[2]

[2] Metrical feet are groups of one to three (usually two) syllables such that one is STRONG (stressed) and the others are WEAK (or unstressed).

c. In many languages, tone is assigned to syllables
d. In many languages, certain sounds can only occur in the ONSET (the part of the syllable before the NUCLEUS) or the CODA (the part of the syllable after the NUCLEUS) of syllables.
   i. In English, the velar nasal /ŋ/ (the <ng> sound) can only occur in codas (*mung* but not *ngum*)
   ii. In Mandarin, all consonants **except** /n/ and /ŋ/ can only occur as onsets (and /ŋ/ can only occur as a coda)

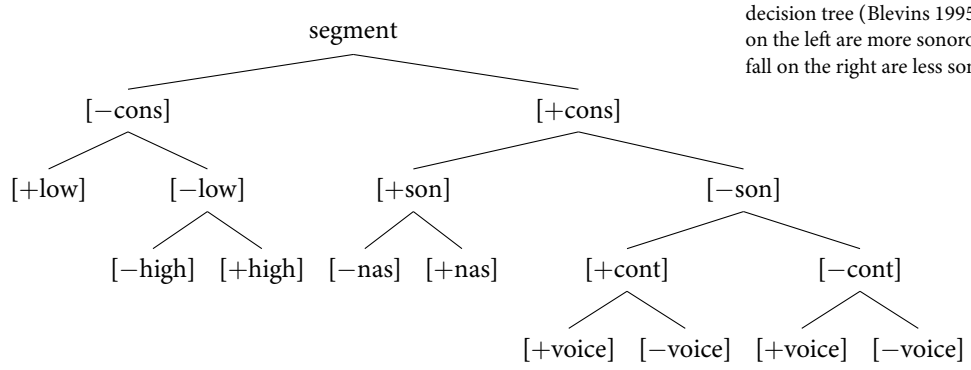
   ,
e. Certain phonological alternations are based on positions in syllables.
   i. In Catalan, plosives in coda position become [−voice].
   ii. In English, plosives in onsets are aspirated (if the syllable is stressed or initial) but coda plosives are never stressed.
f. In many languages, including all modern varieties of Chinese, morphemes correspond roughly to syllables. In many other languages, morpheme boundaries are always at syllable boundaries.
g. Speakers of most languages are able to "clap out" words in the language; these rhythmic units correspond to syllables.

Defining Sonority

Syllables are tied up with the notion of SONORITY. Syllable nuclei must be sonority peaks. However, sonority is a rather abstract notion. Juliette Blevins sought to formalize this notion in terms of articulatory features.[3] She proposed the scale described by the decision tree in Figure 2.

[3] Juliette Blevins. The syllable in phonological theory. 1995. URL `https://api.semanticscholar.org/CorpusID:59673610`

Figure 2: A universal sonority scale as a decision tree (Blevins 1995). Items that fall on the left are more sonorous and those that fall on the right are less sonorous.



A version of this decision tree is implemented in PanPhon, allowing one to automatically calculate sonority values for arbitrary phonemes. These sonority values can, in turn, be used to develop sonority-based syllabification algorithms.

An example of such an algorithm is shown in 1. This algorithm is implemented in `https://github.com/dmort27/syllabiphon` and is described in a recent paper[4].

[4] Ryan Soh-Eun Shim, Kalvin Chang, and David R. Mortensen. Phonotactic complexity across dialects, 2024

Algorithm 1: Algorithm for finding syllable boundaries given a list of sonority scores

function FINDBOUNDARIES(s: list of sonority scores)
    $P \leftarrow [\,]$
    for each $i \in 0...|s| - 1$ **do**
        if $s[i] >= \alpha$ **then**
            push(P, i)
    $b \leftarrow [\text{True} \mid x \in s]$
    push(b, True)
    for each $i \in 0...\text{length}(b[i : -1])$ **do**
        if $i < P[0]$ **or** $i > P[-1]$ **then**
            $b[i] \leftarrow$ False
        if $s[i-1] < s[i]$ **then**
            $b[i] \leftarrow$ False
        else if $i = \text{length}(s) - 1$ **then**
            $b[i] \leftarrow$ False
        if $i > 2$ **and** $s[i-2] = s[i-1] = s[i]$ **then**
            $b[i] \leftarrow$ False
    return b

This algorithm for syllabification performs less well than supervised algorithms based, e.g., on HMMs, but it has the advantage of requiring not training data and no language-specific tuning in order to produce meaningful syllable segmentations.

Miscreants

There are some notable problems with generalizations about sonority. The most egregious malcontents, in this respect, are voiceless fricatives (like English /s/). Consider the following examples:

(3) a. pat /pæt/
    b. pass /pæs/
    c. pats /pæts/
    d. past /pæst/
    e. pasts /pæsts/

If /s/ is more sonorous than /t/, as suggested by Blevins's hierarchy, then we would expect pat, *pass* and *past*, but we would not expect *pats* or *pasts*. We also would not expect *spat* but might expect *psat*.

These misbehaving /s/s only occur at word boundaries. One might address the problem seen here by stetting the sonority value of fricatives at word

boundaries to the minimal value before applying segmentation.

References

Juliette Blevins. The syllable in phonological theory. 1995. URL `https://api.semanticscholar.org/CorpusID:`59673610.

Ryan Soh-Eun Shim, Kalvin Chang, and David R. Mortensen. Phonotactic complexity across dialects, 2024.