Linguistics 051
Proto-Indo-European Language and Society
Proto-Indo-European Sound System

The sound segment inventory of PIE differentiates four important classes of segments based on their *sonority*, i.e. the relative amount of acoustic energy that the sound produces.

1 Obstruents abbreviation: *T*

1	2	3	4	5	*manners*	*places*
p	t	$\hat{k}$	k	k^w	voiceless stop	1: labial
b	d	$\hat{g}$	g	g^w	voiced stop	2: dental
b^h	d^h	$\hat{g}^h$	g^h	g^{hw}	voiced aspirated stop	3: palatal
	s					4: velar
						5: labialized-velar

The stops and *s* are *obstruents*, in which there is sufficient constriction in the oral cavity that air pressure in the oral cavity rises in comparison to the ambient air pressure.

2 Laryngeals abbrevation: *H*

3	4	5	
h_1	h_2	h_3	'laryngeal'

The laryngeals appear to have been somewhere between obstruents and the resonants (below). For most purpose they can be treated as having the same sonority as the obstruents.

3 Resonants abbreviation: *R*

1	2	3	
m	n		nasal
	l		lateral
	r		rhotic
u		i	high vocoid

4 Vowels abbreviation: *V*

6	7		
e	o	mid vowel	6: front
a		low vowel	7: back

In PIE the only proper vowels are *e o a.* These segments are always *syllabic*, that is, for each instance of a vowel there is a syllable for which the vowel is the *nucleus*, or peak of sonority.

As in English, the *resonants* can also become syllable nuclei, *provided that they are not adjacent to a vowel.*

Laryngeals

- The 'laryngeals' are a category of segments whose precise phonetic properties are quite uncertain
- The laryngeals were first reconstructed by Ferdinand de Saussure. He called them *coéfficents sonantiques* and wrote them $*ə_1$ $*ə_2$ $*ə_3$
- In terms of their behavior in syllables and because of the possible 'gaps' in the segment inventory, they seem to fit into the system as fricatives.
- Moreover, when Hittite was deciphered in the 1920s some of the hypothesized laryngeals appeared in Hittite words, spelled with a symbol corresponding to a /x/ like sound in Semitic languages
- This has suggested to some a phonetic realization of /ç x x^w/; others believe some may have been pharyngeals like /ʕ h/
- Some evidence suggests that $*h_3$ was voiced

The laryngeals have 4 principal effects:

a. Coloring: $*h_2$ causes an adjacent *ĕ to become *a
$*h_3$ causes an adjacent *ĕ to become *o

b. Lengthening: $*\mathbf{VH} > *\bar{\mathbf{V}}$ when V and H are *tautosyllabic,* i.e., contained in the same syllable.

c. 'Vocalization': In certain cases where a laryngeal appears at the edge of a word before or after a consonant (i.e. not a vowel), or in the middle of a word between consonants, a 'prop' vowel was introduced to permit the cluster of consonants to be pronounced. This process is called vowel *epenthesis*. Because the laryngeal eventually disappears but the epenthetic vowel normally survives, it *looks* as if the laryngeal has 'turned into'a vowel.

Example: $*ph_2ter$- 'father' > $*p^əh_2ter$- > Gk and L *pater*-

In the 19th century before the discovery and decipherment of Hittite, some Indo-Europeanists believed that PIE had a schwa /ə/ segment, an additional vowel. Eventually it was shown that this schwa — the so-called 'first schwa' or **schwa primum** — was the result of 'vocalized' laryngeals.

Famously, only in *Greek* the laryngeals 'vocalize' differently: $*h_1 > e$, $*h_2 > a$, $*h_3 > o$

d. Tone: A laryngeal can affect the tone or 'accent' of words in Balto-Slavic languages. We will not be looking at this complex problem!

Laryngeal Lengthening and Coloring

- *h_1 did not cause a change in a neighboring *e
- *h_2 caused an adjacent *ĕ to become *a

  *h_2eĝ- 'to lead' > *h_2aĝ-, Gk *ágō*, L *agere* (ACTION)

  *peh$_2$ur- 'fire' > *pah$_2$ur, H *paḫḫur*, Gk *pūr* (PYRE), NE FIRE

  *neu̯eh$_2$- 'make new' > *neu̯ah$_2$-, H *newaḫḫ-* 'id.', Gk *neá-ō* 're-plow'
- *h_3 caused an adjacent *ĕ to become *o

  *$h_3$er- 'eagle' > *$h_3$or- > H *ḫār-as*, Gk *órnis* (ORNITHOLOGY)

  *ĝneh$_3$- 'to know' > *ĝnoh$_3$- > Gk *gi-gnṓ-skō*, R *zna-t′*, NE KNOW
- Note that if a laryngeal follows a vowel in the same syllable, the vowel become long

  *eh_1 > *ē eh_2 > *ah_2 > *ā *eh_3 > *oh_3 > ō

Syllable Structure

Syllables are subgroupings of strings of segments; generally the groups that are formed can be predicted based on the properties the segments have.

The key concept for syllable structure is **sonority**, an approximate measure of the amount of energy required in the production of a type of segment. In PIE the sonority scale is:

most sonorous *least sonorous*

non-high vowel > high vocoid > sonorant > obstruent

It is customary to abbreviate certain categories of segments as follows:

V = vowel (a non-high vowel: *a e o*)
R = high vocoid or sonorant (*l m n r i u*)
T = an obstruent (fricative or stop)
H = a laryngeal (*h_1, *h_2, *h_3)

Sonority Sequencing Principle

The **Sonority Sequencing Principle** requires that:

- syllables contain a **single sonority peak** — the **nucleus**, and
- they **do not fall in sonority before the peak** — in the **onset** — and
- they **do not rise in sonority after the peak** — in the **coda.**

The nucleus and coda together are referred to as the **rhyme** of the syllable.

An unusual property of PIE syllables is that **all segments except stops and [s] can be syllable sonority peaks.**

Example

```
 σ     σ
 |    / | \
 N  O N C
 |  | | |
* r  s ē n
```

[r] is syllabic so it is written *r̥

*r̥sēn ‘male’ > Gk *ársēn*, Skt *r̥sabhá-*

Schematically:

```
      x
x     x  x
x  x  x  x
R  T  V  R
```

There are two **sonority peaks** or **maxima**: the first R and the V. Each of these forms a syllable.

Onset First Principle
When a single segment (in *r̥sēn* it is the *s*) stands between two maxima it joins **the following segment’s syllable**.

This is part of a larger generalization that we can think of as:

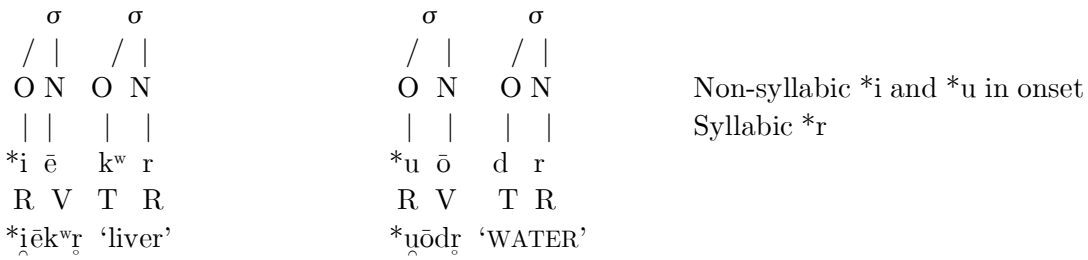
Syllables ‘like’ to have onsets. Onsets are ‘good’.

Syllabic and non-syllabic resonants

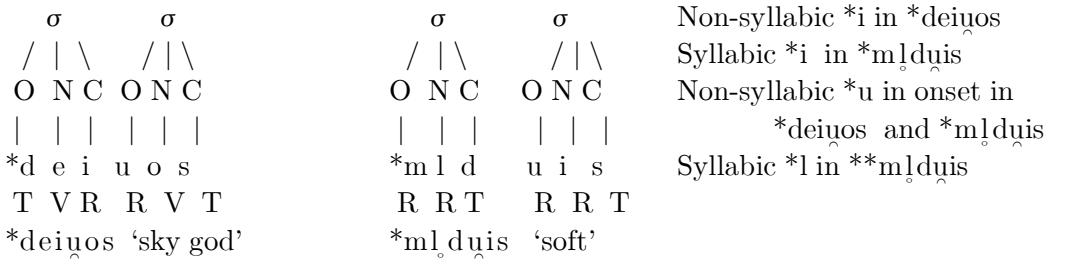
Depending on what the adjacent segments are the resonants can end up being either 'syllabic' (in the syllable nucleus position) or 'non-syllabic' (in the syllable onset or coda).

When *i and *u are non-syllabic they are called **glides.** When they are in the syllable onset they are written either as *i̯ *u̯ or as *y (or *j) and *w. Otherwise they are simply written *i and *u.

Similarly, the sonorants can end up in a syllable nucleus. In this case they are written with a little circle underneath to show that they are 'syllabic': *m̥ *n̥ *l̥ *r̥.



> Gk *hēpar* (HEPATITIS) > H *watar*



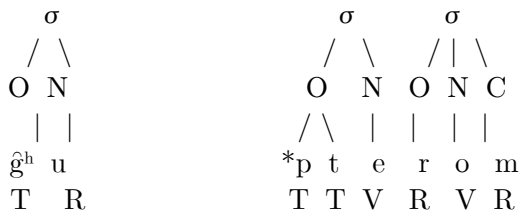
> Gk *sdeús* 'ZEUS', L *deus* > L *mollis* (MOLLIFY)

Notational Summary

segment	onset	nucleus	coda	
*e,*a,*o	—	e,a,o	—	•vowels are always in the nucleus
*u	u̯/w	u	u	• resonants can be either in
*i	i̯/y/j	i	i	the nucleus or in the 'margin'
*m	m	m̥	m	(= onset or coda)
*n	n	n̥	n	
*l	l	l̥	l	
*r	r	r̥	r	
*h	h	—	h	• laryngeals, *s* and stops are
*s	s	—	s	never in the nucleus
*p (stops)	p	—	p	

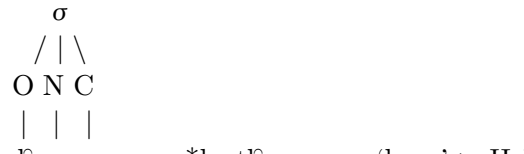
Complex onsets

- An onset that has more than one segment in it is called **complex**.
- A sequence of two or more non-syllabic segments is called a (consonant) **cluster**.
- A complex onset with three consonants usually has [s] as one of the consonants.
- But a complex onset with two consonants can have any two, provided there is no **fall** in sonority:



*mreĝʰu ‘short’
> Gk brakʰús (BRACHY-)

*pteróm ‘wing’
> pterón (PTERO-DACTYL)



*h₂r̥tk̂os ‘bear’ > H *ḫartkas*, Gk *árktos*, L *ursa* (ARCTURUS, URSA MAJOR)

In *h₂r̥tk̂os the *r* is syllabic: it is surrounded by segments of less sonority. So the *r* is a sonority peak and so becomes the nucleus of the first syllable. We write a small circle underneath to indicate that the *r* is syllabic: *r̥*.

*ĝl̥h₃u̯ōs ‘husband’s sister’ > L *glōs,* Grk *gálōs*, Skt *girí-*

The *u* is not a sonority peak since it is adjacent to a more sonorous segment *o*. Thus it becomes the onset of the second syllable. We write a small inverted ‘breve’ mark under an *i* or *u* to show that it is *not* syllabic: *u̯*

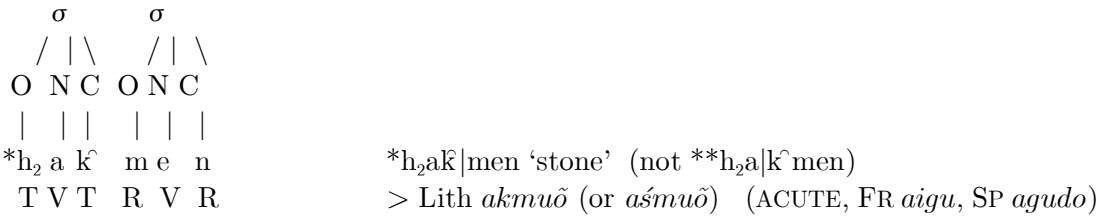
One might ask then why *h$_2$r̥t|k̂os is preferred over *h$_2$r̥|tk̂os or *h$_2$r̥tk̂|os. By the Onset First Principle, *h$_2$r̥tk̂|os is going to be dispreferred since the second syllable has no onset. However, in *h$_2$r̥|tk̂os both syllables have onsets, so what determines which is better? The answer is that:

Complex is 'bad'
Complex codas and onsets are avoided when possible. 'Complex' is 'bad'.

Thus we can see that whenever you have VCV you will syllabify V|CV by Onset First.

When you have VCCV you will normally syllabify VC|CV since Complex is 'bad'.

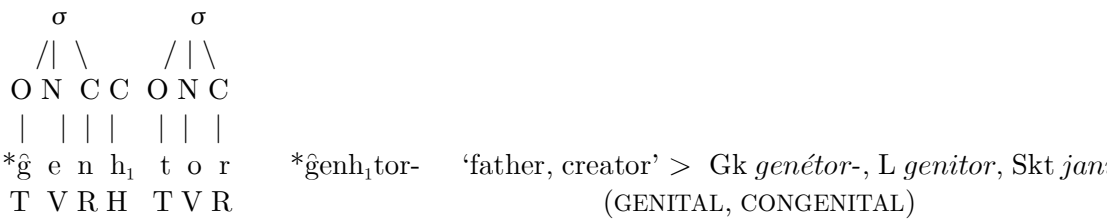
Internal clusters
Whenever a sequence of non-nuclear segments appears between two syllable nuclei, the two non-nuclear segments will split between the two syllables that are formed:



*h$_2$ak̂|men 'stone' (not **h$_2$a|k̂men)
> Lith *akmuõ* (or *aśmuõ*) (ACUTE, FR *aigu*, SP *agudo*)

While all human languages obey Onset First, there is some variation in how CC clusters are divided in different languages. For the purpose of the class we'll assume that in PIE the division is normally VC|CV.

A 'complex' problem:
In the following word we have a laryngeal 'stranded' between a non-syllabic resonant and stop:

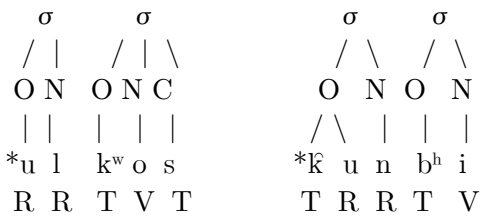


*ĝenh$_1$tor- 'father, creator' > Gk *genétor-*, L *genitor*, Skt *janitár-* (GENITAL, CONGENITAL)

Ultimately the *h$_1$ 'vocalized', giving *e* in Greek, and *i* in Latin and Sanskrit. Presumably the *h$_1$ joined with *n to form a *complex coda.* This is presumably a Sonority Sequencing effect: given the choice between *ĝenh$_1$|tor and *ĝen|h$_1$tor, the former is 'better' since *h$_1$ is less sonorous than *n, and so the first syllable will still fall in sonority, while in *ĝen|h$_1$tor there is certain no rise in sonority in the second syllable. (Effectively, where there is a choice, a laryngeal 'prefers' not to be syllabified as part of a complex onset where the second segment is a stop; sometimes, though, there's no other choice).

'Trapped' sequences RR: an asymmetry

When two resonants are adjacent and the group is neither preceded nor followed a vowel whose syllable one of the resonants can join, we can call the sequence 'trapped'. Here the general rule is that **the first resonant will be the onset and the second the nucleus**:



*u̯l̥kʷos 'WOLF'	*ku̯n̥bʰi 'to the dogs'
> L *lupus*	(*kun- > Gk *kunós* 'dog')
	(CYNIC)
(not **ulkʷos)	(not **kunbʰi)

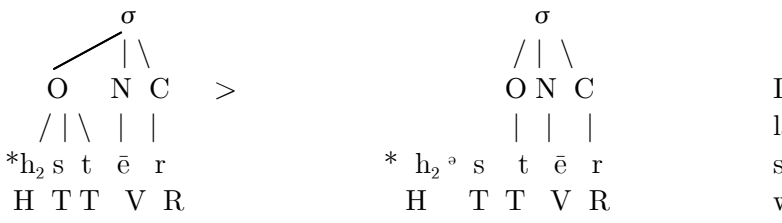
In *u̯l̥kʷos 'wolf' it is clear that Onset First will favor making the *l* the nucleus.

This is not the case for *ku̯n̥bʰi 'to the dogs', however, since the first syllable of *ku̯n̥bʰi actually has a *complex* onset, which it could have avoided if syllabified *kunbʰi with the *u as the nucleus.

There is some evidence that PIE obeyed a special rule favoring RR̥ over R̥R for 'trapped' RR.

Onset clusters beginning with laryngeals

*HT- and *HR- onset clusters are typical situations in which laryngeals vocalized, giving the so-called *schwa primum.* In words in Greek, Armenian and Phrygian in particular, a real vowel is the regular result in these cases.



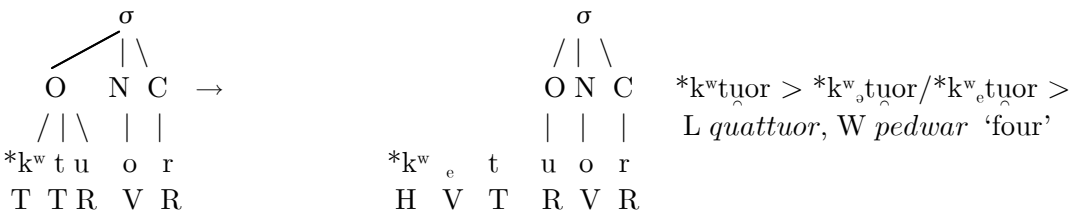
*h₂stēr 'star' >	*h̥₂stēr > Gk astēr 'star'	(cf. STAR)

In all the other IE lgg. the laryngeal remained non-syllabic and was deleted when in the onset.

Schwa secundum ('second schwa')

Onset clusters consisting of stop+stop+sonorant (TTR-) also developed a 'prop' vowel to separate the first two stops in the cluster.

It is customary in IE studies to write the epenthetic vowel as a subscript /$_{e}$/ to show that it may originally have been very short:



Ultimately the schwa primum and the schwa secundum probably arose from similar forces, namely as a resolution to the 'problem' of having too many non-syllabic segments in a row.

Whereas earliest PIE was probably very permissive, allowing large clusters of non-syllabic segments, all the documented languages appear to have developed prop vowels or deleted segments in order to simplify these clusters.

Long vowels
In addition to short vowels, PIE also had long vowels *ē *ō *ū.

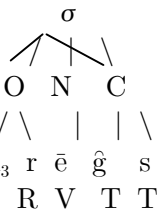
Long vowels that developed in later PIE could either be 'original' (i.e. archaic long vowels) or could be derived from a sequence of a short vowel plus a laryngeal in the syllable coda.

Because in these cases it appears that the laryngeal has been lost and the vowel acquires the duration or 'place' of the lost laryngeal, phonologists call this process **compensatory lengthening.**

For lengthening to occur the vowel and the following laryngeal must be in the same syllable. In this case they are said to be *tautosyllabic* (same-syllable).

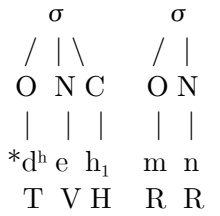
If they are not in the same syllable they are said to be *heterosyllabic.*

original (archaic) long *ē



*h$_3$rēĝ-s (→ *h$_3$rēk̂s)
> L *rēx* 'king' (cf. REGAL)

*ē derived from tautosyllabic *eH



*dʰeh$_1$m̥n → *dʰēm̥n 'thing placed'
Gk (*ana-*)*tʰēma* 'offering' (cf. THEME)

Long high vocoids and sonorants

The evidence for original long high vocoids *ī *ū is less robust than for original long vowels.

Most instances of *ī *ū are probably derived from tautosyllabic *iH, *uH.

Similarly, long sonorants such as *r̥̄ developed from tautosyllabic *RH.

There are a few cases where original long high vocoids are usually postulated:

*mūs 'MOUSE', *nū 'NOW', *û̯īr- 'poison' > L *vīr-* (cf. VIRUS)

Coda Cluster rules

- Two rules of PIE phonology affected clusters of consonants in syllable codas.
- By far the more important of these is:

Szemerényi's Law
A coda cluster *R+s changes to *R, losing the *s, and the preceding nucleus is compensatorily lengthened.

```
            σ
   / /    / | \
 O O   O  N  C
 | |   |  |  | \
*p h₂  t  e  r  s     →     *p h₂  t  ē  r          > 'FATHER', Gk patēr
 T T   T  V  R  T           T T   T  V  R
```

- The second rule is much more restricted in scope, so it is not necessary to memorize it.

A complex coda consisting of *RN or *HN (N = a nasal) is unexpected by the rules of syllabification, since we would expect the nasal to be syllabic. Such cases were affected by:

Stang's Law

In word-final clusters of *RN or *HN (N = a nasal), the *R or *H preceding the nasal was lost, with compensatory lengthening of the preceding syllable nucleus:

*gʷoum expected form: *gʷo̯u̯m̥ (would have > Skt **gāva*)
actual form: *gʷōm > Skt *gām* 'cow' asg.

The typical examples involve a nasal, but some evidence suggests that the rule sometimes applied in the more general cases *RR and *HR.

*ih₂ > *ī in some lgg. (e.g. Skt) but *i̯a in Gk (and sometimes) Tocharian.

*potn-ih₂ 'lady' > Av *paθnī* but Gk *pótnia*

The following three phenomena affect clusters of obstruents.
Usually they are not very important but we will probably see their effects later.

Dental+Velar 'Thorn' Clusters

Dental plus velar clusters gave rise to complicated changes.
For many years it was not clear how to reconstruct this sequence, and it was conventional to write the reconstruction *þ with the Icelandic and Old English letter 'thorn'.

*h$_2$r̥tk̂os 'bear' > Skt *ṛ́kṣas*, Gk *árktos*, L *ursus*
*dʰĝʰom- 'earth' > H [tkan], Luw [-dzgan], Skt *kṣam-*, Gk *kʰtʰon-*,
L *hum-us,* TA *tkaṃ*

In Skt and Gk the segments in the cluster appear to have changed places ('metathesized' or undergone 'metathesis'). In L(atin) the cluster has simplified. The earliest splitting groups Anatolian (Hittite: *H* and Luwian: *Luw*) and Tocharian preserve the original order.

Affrication in Clusters

On the basis of Luwian and comparative evidence it has been suggested that in PIE a dental became 'affricated' when preceding a velar:

*t+k → [ʦk] *d+g → [dzg]

An affricate is a sound produced with two phases of articulation
• stop phase — no air passing out of the mouth
• release phase — air passes out, but the aperture is narrow like a fricative

The first dental in a cluster of two dental stops also became affricated:

*t+t → [ʦt] *d+d → [dzd]

Voicing Agreement in Clusters

In a cluster of two obstruents, the first obstruent acquires the voicing of the second. This process is known as voicing 'assimilation':

*sd → [zd] *ĝ+t → [k̂t]
*dt → [tt] → [ʦt]

*leĝ-tó-s → *lek̂tós > Gk *lektós* 'chosen' (cf. ELECT < L *ē-lectus*)
*nisdo-s → *nizdos > Lith *lìzdas*, OCS *gnězdo* 'NEST'
*h$_1$ed-te- → *h$_1$eʦte- > H *ēzten* [étsten] 'eat!' (imperative pl.)
*h$_1$éd-ti → *h$_1$eʦti > Lith *ėsti*, OCS *jastŭ* 'eats'

Sievers' Law — discovered by *Eduard Sievers* (pronounced 'Zeefers')

Sievers' Law affects the syllabification of resonants in proto-Germanic and in Indo-Iranian. While it is in fact a tremendously important rule, we will probably see very few cases where it applies. It is included here for your reference only.

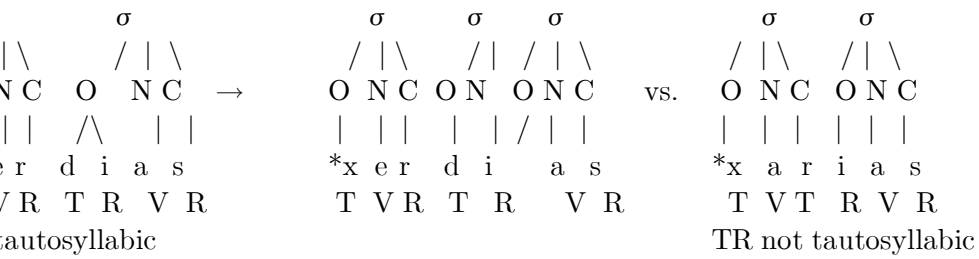
- When a complex onset consists of a consonant plus glide, the glide becomes syllabic (a syllable nucleus) if the preceding syllable is *heavy*, i.e. it has either a coda or a long vowel. Thus:

*VX|C$\underset{\frown}{\text{i}}$V > *VX|Ci|$\underset{\frown}{\text{i}}$V C = consonant, | = σ boundary
*VX|C$\underset{\frown}{\text{u}}$V > *VX|Cu|$\underset{\frown}{\text{u}}$V X = coda or vowel length

Writing j for *i$_{\frown}$, as is traditional for Germanic lgg., we get:

*VX|CjV > *VX|Ci|jV

Proto-Gmc. *xerdjas 'shepherd' > *xerdijas > Goth *hairdeis* where Goth *ei* <*ija
Proto-Gmc. *xarjas does not undergo the Law and becomes Goth *harjis* (the glide *j* remains)



After the *i becomes syllabic the word has two vowels in a row: *xerdias.

When 2 σ nuclei are adjacent (with no intervening onset) it is called **hiatus**. In many languages hiatus is not tolerated: an onset must be inserted.

To 'repair' the hiatus, here the high vocoid syllabifies as **both** the nucleus of the 2nd syllable and as the onset of the final syllable. This is written *ij*.