One basic assumption (a.o.):

Phonotactic constraints act in many languages as word-boundary cues.

Early references:

“[…] I would go further, and say that a ‘word’ is a phonetic entity – that the blank spaces between written words do have phonetic significance.” (Jones 1931 : 154)

“The demarcative function (DF) of phonic elements, consists of isolating each semantic unity in the spoken chain. Demarcative signals can be positive when they indicate the limit of a word, and negative when they indicate that at such point a word boundary does not exist […].” (Anderson 1965 : 185)

Word-Sensitive Phonological Phenomena
(Cross-linguistic evidence)

Vowel Harmony: One feature is mandatorily shared by all vowels of the word:

--- Finnish, Hungarian [FINNO-UGRIC], ...
(Van Der Hulst & Van De Weijer 1995 : 498-499)
**Stress Assignment:**

Stress saturates the word: no words with more than one stress are admitted:

\[ \sim (\sigma', \sigma) \_w \]

Stress falls on a fixed syllable of the word (Kager 1995: 368):
- last: French [ROMANCE], ...
- first: Finnish, Hungarian [FINNO-UGRIC], ...
- penultimate: Indonesian [AUSTRONESIAN] (Soderberg & Olson 2008), Salasaca Quichua [QUECHUAN] (Masaquiza & Marlett 2008), ...
- penultimate if heavy, otherwise antepenultimate: Classical Latin [ITALIC].

Stress falls on a fixed morpheme of the word:
- Portuguese [ROMANCE] (Mateus et al. 2003): Nouns: on the last vowel of the root; Verbs: on the Class Marker (Past Tense forms), on the last vowel of the root (Present Tense forms), on the Tense Morpheme (Future Tense forms).

**Prohibition of Segment(s)/Clusters in Word-Initial Position:**

(“negative demarcatrve signals” – Anderson 1965)

A given segment/cluster can never occur at word beginnings:

\[ \sim (\#S_{\text{segment}}) \wedge \sim (\#C_{\text{cluster}}) \]

- Portuguese [ROMANCE]: disallows [n], [k] and [r] word-initially: \( \sim (#k), \sim (#n), \sim (#r) \).
- Kabiye ([VOLTAIC] (Padayodi 2008): disallows voiced obstruents word-initially: \( \sim (#\text{VoicedObstr}) \).
- Nepali [INDO-ARYAN] (Khatiwada 2009): disallows word-initial clusters, unless C2 is a rhotic or a glide: \( ([C_1C_2] \wedge [C_2 \neq (R \vee G)]) \rightarrow \sim (#C_1C_2) \).
- Tamil [DRAVIDIAN] (Kean 2004): disallows retroflex consonants word-initially: \( \sim (#\text{Retroflex}) \).

**Restriction of Segment(s)/Clusters to Word-Initial Position Only:**

(“positive demarcatrve signals” – Anderson 1965)

A given segment/cluster occurs mandatorily at word beginnings:

\( (S_{\text{segment}} \vee C_{\text{cluster}}) \rightarrow ([#S_{\text{segment}}] \vee (#C_{\text{cluster}})) \)

- Yakima Sahaptin [PENUTIAN] (Hargus & Beavert 2006): CCV is admitted word-initially only: (CCV) \( \rightarrow (#CCV) \)
**Salasaca Quichua** [QUECHUAN] (Masaquiza & Marlett 2008): aspirated stops are allowed word-initially only: (AspStop) \rightarrow (#AspStop)

**Prohibition of Segment(s)/Clusters in Word-Final Position:**

("negative demarcative signals" – Anderson 1965)

A given segment/cluster can never occur at word endings:

\[ \sim(\text{Segment} #) \land \sim(\text{Cluster} #) \]

**Indonesian** [AUSTRONESIAN] (Soderberg & Olson 2008), **Ibibio** [NIGER] (Urwa 2004), ...: disallow /b/, /d/, /g/ word-finally: \[C=(b \lor d \lor g)\] \rightarrow \sim(\text{C#})

**Restriction of Segment(s)/Clusters to Word-Final Position Only:**

("positive demarcative signals" – Anderson 1965)

A given segment/cluster occurs mandatorily at word endings:

\[(\text{Segment} \lor \text{Cluster}) \rightarrow [(\text{Segment} \lor \text{Cluster}) \#] \]

**Nepali** [INDO-ARYAN] (Khatiwada 2009): the velar nasal always occurs word-finally: \[C=\eta\] \rightarrow \sim(\text{C#})

**Restriction of Occurrence of Certain Segments/Clusters at Word-Endings:**

Segments/clusters that are the only admitted word-finally (though they can occur in other positions as well – see "onset/coda asymmetry").

\[(\text{C#}) \rightarrow (C \in A) \]

\[ (A=\text{Subset of L consonant inventory}) \]

**Gayo** [AUSTRONESIAN] (Eades & Hajek 2006): Fricatives, nonpalatal nasals, rhotics, laterals and voiceless stops are the only consonants admitted word-finally.

\[[C=\eta] \land [A=\{\text{Fric, NonPalNas, R, Lat, VLessStops}\}] \]

But:

- some evidence of the existence of word-final codas different from word-medial codas:

Peninsular Spanish: admits /d/-filled codas word-finally (Navarro Tomás 1926: 99-100; Quilis 1993: 204-205):
«sed» ‘thirst’,
«huésped» ‘host’,
«césped» ‘grass’,
«juventud» ‘youth’.

Catalan: admits /d3/-filled codas word-finally (even though they are often phonetically deleted; http://en.wikipedia.org/wiki/Catalan_phonology).

THE CASE OF EUROPEAN PORTUGUESE

European Portuguese phonology:
highly restrictive as far as coda-filling is concerned:

- filled codas are less frequent than empty codas
- no complex codas
- only /l/, /ɾ/ and /S/ are admitted as coda-fillers (see, e.g. Mateus & Andrade 2000)

But:

Word-finally,
- ‘extra-heavy rhymes’
- codas filled by consonants different from /l/, /ɾ/ or /S/.
(see, e.g., Veloso 2008, 2009, forthcoming)

Extra-heavy rhymes admitted word-finally only in European Portuguese:

<table>
<thead>
<tr>
<th>/VGN/#</th>
<th>/VGNsp/#</th>
<th>/VGNspred/ #</th>
</tr>
</thead>
<tbody>
<tr>
<td>« pão» ‘bread’ [pẽ̃u]</td>
<td>« mãos » ‘hands’</td>
<td>« Guimarães » (place-name)</td>
</tr>
<tr>
<td>« ontem » ‘yesterday’ [õمست]</td>
<td>« irmãos » ‘brothers’ [iɾ’mẽ̃uʃ]</td>
<td>[giẽmoʃ]</td>
</tr>
<tr>
<td>« homem » ‘man’ [õmoʃ]</td>
<td>« alemães » ‘German [plural]’</td>
<td>« Simões » (person-name)</td>
</tr>
<tr>
<td>« ruim » ‘bad’ [ɾũ̃]</td>
<td>[eĩ’mẽ̃uʃ]</td>
<td>[si’mõ̃oʃ]</td>
</tr>
</tbody>
</table>

Unattested word-medially: *[pẽ̃u.tu], *[mẽ̃uʃ.tru]

Word-final codas filled by consonants different from /l/, /ɾ/ and /S/:

Segmental /n/:
- « gérmen » ‘germ’, [ʒẽr̃ẽm] |
- « plâncion » ‘plankton’, [pl̃ĩk̃õ̃n] |
  Unattested word-medially: *[kẽntu], *[sĩ̃r’mẽñ.gu]

<table>
<thead>
<tr>
<th>/ks/#</th>
<th>/ps/#</th>
</tr>
</thead>
<tbody>
<tr>
<td>« tórax » ‘thorax’ [tɔʁ̃ek̃]</td>
<td>« fórceps » ‘forceps’</td>
</tr>
<tr>
<td>[fɔɾseps]</td>
<td></td>
</tr>
<tr>
<td>« córtex » ‘cortex’ [kɔɾtek̃]</td>
<td>« bíceps » ‘biceps’ [biseps]</td>
</tr>
</tbody>
</table>
  Unattested word-medially: *[teks.tu], *[laps.tu]

“PROSODIC TOLERANCE
OF THE WORD RIGHT BOUNDARY”
(PTWRB)
(Veloso forthcoming)

Declarative Phonology-Based Formalizations
of Possible “PTWRB Phenomena”
and Other Word-Boundary- Sensitive
Phonotactic Constraints in Romance

P.1: Prohibition of word-initial /ʎ/, /ɲ/ and /ɾ/ in European Portuguese
[Seg = (ʎ ∨ ɲ ∨ r)] → ~[#init Ʌ Seg]

P.2: PTWRB in European Portuguese
{Seq = [(n.) ∨ ((k ∨ p)s) ∨ (VGN) ∨ (VGNS)]} → [{Seq Ʌ #Fin} ∧ (. = #Fin)]

S.1: Word-final /d/-codas in Pensinsular Spanish
[Seq=(d.)] → [(Seq Ʌ #Fin) ∧ (.=#Fin)]

C.1: Word-final /dʒ/-codas in Catalan
[Seq=(dʒ.)] → [(Seq Ʌ #Fin) ∧ (.=#Fin)]

Key
P = (European) Portuguese  Init = Initial
S = Spanish  Fin = Final
C = Catalan  Ʌ = precedes immediately
Seg=Segment  # = Word boundary
Seq=Segment combination
. = Syllable boundary
V, G, N, S = Vowel, Glide,
Nasal, Palatal Fricative
C = Consonant

Final Remarks
1 – Some phonotactic constraints can be accepted as word-boundary cues at least in some languages.

2 – Apart from the descriptive interest of this regularity, its relevance is twofold:
   a) it can explain how hearers may identify word-boundaries in speech processing tasks;
   b) it may be useful for the development of automatic tools for word demarcation within speech continua.

3 – It must be added, though, that in languages where PTRWB is found, it is a sufficient but unnecessary condition for word boundary identification.

4 – This kind of linguistic data offer us an extra amount of evidence favouring the word as a linguistic domain/unit and the necessity of including phonological aspects among the “wordhood criteria”.

5 – Declarative Phonology-based formalisms (Scobie et al. 1996; Angoujard 2006) offer adequate descriptions of all relevant, surface-observable phonotactic regularities of the lexicon of a given language.

REFERENCES


