Automatic detection of syllable boundaries in spontaneous speech

Brigitte Bigi
Christine Meunier, Irina Nesterenko, Roxane Bertrand

Laboratoire Parole et Langage
LREC - May 2010
Main goal

- Automatic detection of syllable boundaries
- Development of a Rule-Based System (RBS):
  - for automatic syllabification of phonemes’ strings
  - of the size greater than a graphic word
  - Example:

  Phonemes  i l e k s p l i k e p a v r e m ā s k i i a v e d ā

- Application to conversational speech
  - 8 French informal dialogues
The syllabification process is based on 2 main principles:

1. A syllable contains a vowel, and only one.
2. A pause is a syllable boundary.

These two principles bring the problem to find the boundaries between two vowels.

Phonemes: il e k s p l i k e p A v r e m Å s k i j A v e d Å
Related works (1)

- `syllabation.awk`, made by C. Pallier
  - GPL
  - To segment phonetized words into syllables
  - Phonemes are grouped into 4 classes:
    - vowels, glides, liquids and other consonants
  - 9 segmentation rules are established
    - to find the boundary between 2 vowels
    - by using the classes or
    - by using the phonemes directly in cases a class is not relevant
  - Successfully applied to lexical databases: Brulex and Lexique
M. Adda-Decker, P. Boula de Mareüil, G. Adda, L. Lamel
"Investigating syllabic structures and their variation in spontaneous French", SpeechCom, 2005

- Software not available
- Part of graphon+
- To segment phonetized spoken French
- Phonemes are grouped into 4 classes:
  - vowels, glides, liquids, other consonants
- 13 segmentation rules are established
  - to find the boundary between 2 vowels
  - by using the classes or
  - by using the phonemes directly in cases a class is not relevant
syllabify2.praat, made by J.P. Goldman

- GPL
- Part of EasyAlign software
- To segment phonetized spoken French
- Phonemes are grouped into 6 classes:
  - silence, vowels, glides, liquids, [p t k b d g f v] and [s z z m n ñ ñ]
- About 60 segmentation rules are established
  - to find the boundary using classes
  - not specifically between vowels
Group phonemes into 6 classes

V - Vowels: i e æ a ø o u y ø œ ø ɛ ã ů ø
G - Glides: j ɥ w
L - Liquids: l r
O - Occlusives: p t k b d g
F - Fricatives: s z ʒ ʃ f v
N - Nasals: m n ɳ ɲ

Unlike other systems, we divide consonants into 3 classes: O, F, N.

<table>
<thead>
<tr>
<th>Phonemes</th>
<th>i l e k s p l i k e p a v r e m ā s k i j a v e d ţ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classes</td>
<td>V G V O F O L V O V O V F L V N V F O V G V F V O V</td>
</tr>
</tbody>
</table>
The letter X to mention one of G, L, O, N or F.

<table>
<thead>
<tr>
<th>Observed sequence</th>
<th>Segmentation rule</th>
<th>Examples (French)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 VV</td>
<td>V.V</td>
<td>poëte : po.ët, il y a un : i.a.œ</td>
</tr>
<tr>
<td>2 VXV</td>
<td>V.XV</td>
<td>limité : li.mi.te, et donc on : e.dõ.kõ</td>
</tr>
<tr>
<td>3 VXXV</td>
<td>VX.XV</td>
<td>jardin : zâr.dê, comme ça : kom.sa</td>
</tr>
<tr>
<td>4 VXXXXV</td>
<td>VX.XXV</td>
<td>avec moi : a.vek.mwa</td>
</tr>
<tr>
<td>5 VXXXXXV</td>
<td>VX.XXXV</td>
<td>il se présentait : il.spêrezâ.te</td>
</tr>
<tr>
<td>6 VXXXXXXV</td>
<td>VXX.XXXV</td>
<td>alors je crois : a.lorž.krwa</td>
</tr>
</tbody>
</table>
## Exception rules

<table>
<thead>
<tr>
<th>Observed sequence</th>
<th>Segmentation rule</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>VXGV</td>
<td>V.XGV</td>
<td>baignoire : be.nwar, spéciaux : spe.sjo</td>
</tr>
<tr>
<td>VFLV</td>
<td>V.FLV</td>
<td>découvrez : de.ku.vra,</td>
</tr>
<tr>
<td>VOLV</td>
<td>V.OLV</td>
<td>il trouve : i.truv, mais de la : me.dla</td>
</tr>
<tr>
<td>VFLGV</td>
<td>V.FLGV</td>
<td>effroyable : ef.rwa.jabl</td>
</tr>
<tr>
<td>VOLGV</td>
<td>V.OLGV</td>
<td>incroyable : ë.krwa.jabl</td>
</tr>
<tr>
<td>VOLOV</td>
<td>VOL.OV</td>
<td>connaître tu : ko.netr.ty</td>
</tr>
</tbody>
</table>
The rules we propose follow usual phonological statements for most of the corpus. Our aim is not to propose a true set of syllabification rules for French, but to provide an acceptable syllabification for the most part of spontaneous speech corpus.

<table>
<thead>
<tr>
<th>Transcription</th>
<th>il expliquait pas vraiment ce qu’il y avait dedans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonemes</td>
<td>i l e k s p l i k e p a v r e m ā s k i j a v e d ā</td>
</tr>
<tr>
<td>Classes</td>
<td>v g v o f o l v o v o v f l v n v f o v v v v f v o v</td>
</tr>
<tr>
<td>Syllables Auto</td>
<td>i . lek . spli . ke . pa . vre . mā . ski . ja . ve . dā</td>
</tr>
<tr>
<td>Syllables Expert1</td>
<td>i . lek . spli . ke . pa . vre . mā . ski . ja . ve . dā</td>
</tr>
<tr>
<td>Syllables Expert2</td>
<td>i . leks . pli . ke . pa . vre . mā . ski . ja . ve . dā</td>
</tr>
</tbody>
</table>
The LPL-Syllabeur Tool

- Implemented in java 1.6 and tested under linux and windows®
- GPL
- Input and Output in a TextGrid Praat file
- A configuration file that the user can change as needed to specify
  - phonemes and classes:
    PHONCLASS e V
    PHONCLASS p 0
  - general and exception rules:
    GENRULE VXXV 1
    EXCRULE VFLV 0
  - some phoneme sequences and a boundary shift to apply:
    OTHRULE ANY p s k -2
The LPL-Syllabeur: French or English GUI

**Description file of phonemes, classes and rules:**

- descr/French-CID.txt

**Phonemes Praat file:**

- examples/BX.TextGrid

**Run syllabification**

---

1. Loading description file
2. Loading phoneme file
3. Rules-based Syllabification
4. Save syllables
5. Save statistics
6. Syllabification finished

**Exit**
Corpus description

- The CID - Corpus of Interactional Data
- Audio-video recording of 8 hours of spontaneous French dialogues
- Each dialogue involves two participants (spoke very freely)
- Phonetization from the transcription

An Enriched Orthographic Transcription, which includes:
- Elision, the omission of one or more sounds
  - \( j'\text{ai on } a j'\text{ai p- (en)fin } j'\text{ai trouvé l(e) meilleur moyen } c'(é)tait d(e) loger chez des amis \)
- Particular phonetic realisations
  - \[elle, ] dormait
  - faire des \[stats, stateu\]
- The rate of elision and particular realisation is about 17%
Syllabification of the CID

- 139751 vowels = syllables
- Syllables structures are obtained only after applying segmentation rules between 2 vowels:
  - CV 60.70%
  - V 12.95%
  - CVC 11.46%
  - CCV 10.67%
  - CCVC 1.83%
  - VC 1.37%
  - others are less than 1%

⇒ a correct distribution for French
Syllabification Evaluation

- The test corpus is 1.6% of the CID
  - about 7 minutes of a dialogue
  - 2068 syllables
- The test corpus was manually segmented by two experts
  - a syllable agreement rate of 97.77% (23 boundary mismatches)
- Number of boundary mismatches and syllable difference rate:

<table>
<thead>
<tr>
<th></th>
<th><code>syllabation.awk</code></th>
<th><code>graphon+</code></th>
<th><code>syllabify2.praat</code></th>
<th>LPL-Syllabeur</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(1)</strong> Expert 1</td>
<td>74</td>
<td>80</td>
<td>67</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>7.16%</td>
<td>7.74%</td>
<td>6.48%</td>
<td>4.16%</td>
</tr>
<tr>
<td><strong>(2)</strong> Expert 2</td>
<td>84</td>
<td>85</td>
<td>75</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>8.12%</td>
<td>8.22%</td>
<td>7.25%</td>
<td>5.13%</td>
</tr>
</tbody>
</table>
### Examples

<table>
<thead>
<tr>
<th>Transcription</th>
<th>et donc on mange sur la baignoire donc c'est c'est ça</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonemes</td>
<td>ed₃ k₅ m₃ s y r l a b e n w a r d₃ k s e s e s a</td>
</tr>
<tr>
<td>Classes</td>
<td>V O V O V N V F F V L L V O V N G V L O V O F V F V V</td>
</tr>
<tr>
<td>Syllables (Auto &amp; Experts)</td>
<td>e . d₃ . k₅ . m₃ . s y r . l a . b e . n w a r . d₃ k . s e . s e . s a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transcription</th>
<th>non dans les parcs c'est un peu limité</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonemes</td>
<td>n ₅ d ₉ a l e p a r k s e t ₉ p ε l i m i t e</td>
</tr>
<tr>
<td>Classes</td>
<td>N V O V L V O V L O F V O V O V L V N V O V</td>
</tr>
<tr>
<td>Syllables Auto</td>
<td>n₅ . d₉ . l e . p a r k . k s e . t ₉ ε . p ₑ . l i . m i . t e</td>
</tr>
<tr>
<td>Syllables Experts</td>
<td>n₅ . d₉ . l e . p a r k . s e . t ₉ ε . p ₑ . l i . m i . t e</td>
</tr>
</tbody>
</table>
Mismatches Experts/Automatic examples

- **Transcription**: *offre le*
  - Syllables expert1 and expert2: zof . Iə
  - Syllables auto: zo . flə

- **Transcription**: *comme une*
  - Syllables expert1 and expert2: kom . yn
  - Syllables auto: ko . myn
  - Syllables expert1 and expert2: reks . me
  - Syllables auto: rek . sme

- **Transcription**: *glaces comme*
  - Syllables expert1 and expert2: glas . kom
  - Syllables auto: glə . skom

Experts are influenced by lexical boundaries
In most of the cases, mismatches between the automatic syllabification and the experts’ syllabification concern ambiguous boundaries for which experts propose variable syllabification.

- **Transcription**: *retrouver les*
  - Syllables expert1: αrt . ru
  - Syllables expert2: α . rtru
  - Syllables auto: αr . tru

- **Transcription**: *va se faire*
  - Syllables expert1 and auto: vαs . fεr
  - Syllables expert2: vα . sfer

- **Transcription**: *pas le truc*
  - Syllables expert1 and auto: pαl . tryk
  - Syllables expert2: pα . ltryk

When there are sequences of more than 2 consonants (not quite frequent in French, though often resulting from reduction phenomena), expert2 pays more attention to the lexical material, while expert1 and our algorithm favours a more balanced structure and respects the sonority principle.
A rule-based phoneme to syllable segmentation system

Compared with current state-of-the-art systems, the advantages of the *LPL-Syllabeur* are that:

1. it is made with a small number of simple rules for the syllabification of a spontaneous French corpus in a friendly dialogue context
2. the tool uses an object-oriented language, under GPL license
3. it is very easy to adapt to a specific corpus by adding or modifying rules, phoneme encoding or phoneme classes, by the way of a new configuration file