Automatic detection of syllable boundaries in spontaneous speech

Brigitte Bigi[1], Christine Meunier[1], Irina Nesterenko[2], Roxane Bertrand[1]

Laboratoire Parole et Langage[1],
CNRS & Aix-Marseille Universités,
5, avenue Pasteur,
13100 Aix-en-Provence France,

Institut National des Langues[2]
et Civilisations Orientales
2 rue de Lille
75343 Paris cedex 07 France

e-mail: brigitte.bigi@lpl-aix.fr,christine.meunier@lpl-aix.fr,irina.nesterenko@inalco.fr,roxane.bertrand@lpl-aix.fr

Abstract

This paper presents the outline and performance of an automatic syllable boundary detection system. The syllabification of phonemes is performed with a rule-based system, implemented in a Java program. Phonemes are categorized into 6 classes and specific rules are developed to deal with a French spontaneous speech corpus. Moreover, the proposed phonemes, classes and rules are listed in an external configuration file of the tool (under GPL licence). Finally, performances are evaluated and compared to state-of-the-art systems and show significant improvements.

1. Introduction

In this paper we present an approach to automatic detection of syllable boundaries. We report on the development of a Rule-Based System (RBS) for automatic syllabification of phonemes' strings of the size greater than a graphic word and in application to the conversational speech (our corpus consists of 8 French informal dialogues).

A syllable is a universally recognised linguistic unit. One argument in favour of this unit is the intuition of language users who are able to determine the number of syllables in a word, to enumerate them and to change their order (cf. language games). Linguists appeal to this unit to explain many speech production and speech perception phenomena (see, for example, work on tonal alignment). At the same time, definitions of the syllable in different fields of linguistics show remarkable discrepancies. Thus, in both phonetics and phonology, the syllable is defined as a phonological unit which organises segmental melodies in terms of sonority: first, the sonority of a sound is defined as its inherent loudness, holding factors like pitch and duration constant (Crystal, 2003); next, a syllable is defined as a sequence of speech sounds having a maximum (or peak) of inherent sonority between two sonority minima. Other scholars insist on the role of syllable in accounts of phonotactic constraints operating in a particular language (Pulgram, 1970).

In addition, there is a structural approach to the syllable in which the focus is on syllable-internal structure and sub-syllabic constituents: thus within the syllable, a vowel, which corresponds to the maximum of sonority, is called the syllabic nucleus, the consonants (if any) preceding the nucleus form the syllabic onset, and the consonants (if any) following the nucleus form its coda. These sub-syllabic constituents intervene in the analysis of different phonological phenomena (allophone distributions) and in the processing of the prosodic and temporal organisation of speech. There are a number of theories of syllabification as many scholars have sought general principles and constraints on the syllabic division of a string of phonemes. Thus the Legality Principle (Goslin and Frauenfelder, 2001) constrains the segments that can begin and end a syllable to those that appear at the beginning and end of the words. On the other hand, according to the Maximal Onset Principle, one should extend a syllable's onset at the expense of the preceding syllable's coda whenever it is legal to do so (Kahn, 1976).

The question then arises of how to obtain an acceptable syllabification for a particular language and for a specific corpus (a list of words, a written text or an oral corpus of more or less casual speech). Currently, most state-of-the-art approaches defines syllabification as the process of dividing a word into its constituent syllables. There are two broad approaches to the problem of the automatic syllabification of words: a rule-based approach and a data-driven approach. The rule-based method effectively embodies some theoretical position regarding the syllable, whereas the data-driven paradigm tries to infer new syllabifications from examples syllabified by human experts. Rule-based methods face possible analyses then the rule-based system faces the problem of disambiguation. Probabilistic methods, however, yield the most probable analysis according to the training corpus. (Marchand and Damper, 2007; Marchand et al., 2009) show that rule-based methods are relatively ineffective for orthographic syllabification in English. On the other hand, a few data-driven syllabification systems with a high level of performance currently exist (Bartlett et al., 2008; Bartlett et al., 2009). Unfortunately, to cope with statistical syllabification of words, these techniques are effective only if a syllabified corpus exists for the training.

More difficulties arise when the task is the syllabification of spoken data and casual speech. Phonetization and syllabification of spoken corpora is performed by using word transcription, as for example for Castilian Spanish where a Context Free Grammar is used (Sandoval et al., 2008). First, the orthographic representation of a word is translated into its phonemic transcription. Next, either syllables are extracted from a dictionary or syllabic boundaries are found by applying a set of rules. These approaches do not take into account elision and reduction phenomena which are typical of spoken speech. (Adda-Decker et al., 2005) estimates that a deletion rate of about 15% is observed, while comparing syllabification of an aligned phonemic string and that based on word division. This is mainly due to the fact that in continuous speech there is the process of resyllabification across word boundaries. We assume that it will be even more the case in our French spontaneous speech corpus (see section 2).

Furthermore, a complete study of syllabification of a French radio corpus is proposed in (Adda-Decker et al., 2005). Syllabification is performed with *graphon+* (Boula de Mareüil, 1997). Phonemes are grouped into 4 classes: vowels, glides, liquids and other consonants. A set of 13 rules is proposed to find the boundary between 2 vowels depending on the number of phonemes and their classes. Two more systems were developed earlier for the syllabification of French and are freely available. The first one is called *syllabation.awk*, described in (Pallier, 1999). It is a gawk script to segment phonetized words into syllables. Phonemes are also grouped into 4 classes: vowels, glides, liquids and other consonants. Nine segmentation rules are established by using the classes or the phonemes directly in cases a class is not relevant. This system has been successfully applied to lexical databases such as Brulex and Lexique(new, 2004). The second freely available system is also a rule-based system described in (Goldman, 2007); *syllabify2.praat* is part of the *EasyAlign* software. It is a Praat script (Boersma and Weenink, 2009), developed for spoken French. Five phoneme classes are defined as: vowels, glides, liquids, [p t k b d g f v] and [s ʃ z ʒ m n ŋ ɲ]. This system uses a large number of rules (about 60) to examine different specific cases, including the silence.

With respect to these already existing systems, the novel aspect of the work reported in this paper is as follows:

- to propose a generic RBS tool to identify syllabic segments from phonemes ;
- to propose relevant rules in the particular context of French spontaneous speech.

Previous work (Blache et al., 2008) presents the corpus and several tools developed for its processing. The corpus is an audio-visual recording of spontaneous spoken French, described in section 2.Our syllabification system is described in section 3:it relies on two principles which frame the problem as a search for the boundary between two vowels. Phonemes are first grouped into relevant classes and rules are then divided in two categories: general rules, that give the boundary between 2 vowels depending on the number of consonants between them, and exception rules operating on the specific sequences found between the two vowels. Section 4. describes the tool we developed. The key part of this tool concerns its configuration file, which is separated to the implementation and allows users to define their own phoneme encoding, their own phoneme classes and all the rules. This choice allows the tool to be adapted to other syllabification tasks or other languages without changing any of the implementation (only the external ASCII configuration file). We discuss the evaluation of our syllabification algorithm in section 5:the performance of our system on a corpus of conversational speech is compared to the three other syllabification systems for French.

2. Spontaneous spoken French corpus

The system described in this paper has been used for the labelling of the CID - Corpus of Interactional Data (Bertrand et al., 2008; Blache et al., 2008). The CID is an audio-video recording of 8 hours of spontaneous French dialogues (1 hour of recording per session). Each dialogue involves two participants of the same gender. One of the following two topics of conversation was suggested to participants: conflicts in their professional environment or funny situations in which participants may have found themselves. These instructions were not exhaustive and participants often spoke very freely about other topics, in a conversational speaking style. The corpus includes data recorded in anechoic room and containing 120,000 words.

One of the characteristics of spontaneous speech is an important gap between a word's phonological form and its phonetic realisations. Specific realisation due to elision or reduction processes (for example *je suis* pronounced as [ʃɥi], *je ne sais pas* as [ʃepɑ]) are frequent in spontaneous data. Some of these instances can be extracted from a lexicon of systematic variants but it will not cover all the observed realisations. Spontaneous data like the CID corpus also present other types of phenomena such as non-standard elisions, substitutions or addition of phonemes (see (Bertrand et al., 2008) for details) which intervene in the automatic detection of syllable boundaries.

The tool we deal with is the phonetizer, which produces the list of phonemes. It is a rule-based system which uses the hand-made enriched orthographic transcription of the signal (TOE). The corpus was orthographically transcribed following transcription guidelines in line with the French GARS conventions (Blanche-Benveniste, 1990). Transcribers were asked to provide an enriched orthographic transcription, which includes, for example, manually annotating non-standard events such as: mispronunciations, truncated words, some liaisons, elisions, laughs, etc. Some of these specificities have a direct consequence on the phonetisation procedure and so on the syllabification:

- Elision is the omission of one or more sounds (such as a vowel, a consonant, or a whole syllable), producing a result that is easier for the speaker to pronounce. Non-standards elisions are explicit in the TOE, manually annotated by parenthesis of the omitted sounds. For example:

  (1) *j'ai on a j'ai p- (en)fin j'ai trouvé l(e) meilleur moyen c'(é)tait d(e) loger chez des amis* 'I've we've I've - well I found the best way was to live in friends' apartment'

  Consequently, the phonetizer will not produce phonemes for elision in the words *enfin*, *le*, etc. Another word frequently produced with elision is *parce que* phonetized as /pɑsk/ or even /psk/ instead of /pɑʀsk/.

- Transcribers also mentioned particular phonetic realisations by using brackets, as the pronunciation of specific words, pronounced schwa, etc. As for example:

(2) *[elle,] dormait* 'She slept'

(3) *du [movetrack, mouvtrac] ouais de de l'[EMA, euma]* 'of movetrack yeah of of EMA'

(4) *faire des [stats, stateu]* 'to do stats'

The "enrichment" rate of elision and particular realisation in the TOE, defined by the ratio of the number of specific phonetic achievements over the number of phonetic tokens is 17%. This high rate confirms our preliminary choice of a manual specification. Based on the phonetisation of this specific corpus, we started the overall process of developing an automatic syllable segmentation system.

3. Proposed Rule-based System (RBS)

The problem we deal with is the syllabification of a phoneme sequences. This means the system copes with entries as for example:
/n ɔ̃ d ɑ̃ l e p ɑ ʀ k s e t œ̃ p ə l i m i t e/
and produces an output with syllable-marks.

Our RBS phoneme-to-syllable segmentation system is based on 2 main principles:

1. a syllable contains a vowel, and only one.
2. a pause is a syllable boundary. The CID corpus was first automatically segmented into inter-pausal units (IPU) delimited by silent pauses of 200 ms and more. Human transcribers marked shorter pauses they perceived. In both cases, a pause signals a syllable boundary.

These two principles focus the problem on the task of finding a syllabic boundary between two vowels.

As in state-of-the-art systems, we propose grouping phonemes into classes and establishing rules dealing with these classes. In our case, we choose to define general rules followed by a small number of exceptions. Consequently, the identification of relevant classes is important for such a system. We propose the following classes:

V - Vowels: i e ɛ a ɑ ɔ o u y ø œ ə ɛ̃ ɑ̃ ɔ̃ œ̃

G - Glides: j ɥ w

L - Liquids: l ʀ

O - Occlusives: p t k b d g

F - Fricatives: s z ʃ ʒ f v

N - Nasals: m n ŋ ɲ

Uppercase letters indicate the abbreviations we will use throughout this paper. We will also use the letter X to mention one of G, L, O, N or F (i.e., a non-vowel phoneme).

Unlike current approaches, we have chosen to divide consonants into 3 classes: O, F and N. This is a very important choice that largely reduces the complexity. Defining fewer classes of consonants either leads to a poor syllabification performance or requires the specification of a large number of exceptions. At this stage of the development, the system does not deal with entries as:
/n ɔ̃ d ɑ̃ l e p ɑ ʀ k s e t œ̃ p ə l i m i t e/
but with:
N V O V L V O V L O F V O V O V L V N V O V.
In this case, the system will place a boundary in: V O V, V L V, V O V, V L O F V, etc.

The general rules developed are listed in Table 1 and exception rules in Table 2.

These general rules are motivated as follows:

- In accordance with principle 1 listed above, there is only one vowel per syllable;
- The second rule reflects the claimed universal tendency to favour open syllables, so that the intervocalic consonant is assigned to the second syllable;
- Rules 4, 5 and 6 satisfy the general Maximum onset principle which specifies that in an intervocalic consonantal cluster as many consonants as possible should be syllabified into the onset of the second syllable rather than the coda of the first syllable;
- The third rule should be regarded in the context of exception rules 1, 2 and 3 (table 2), all of them dealing with two-consonant clusters. Taken together, these rules prove that for two-consonant clusters the Maximum onset principle should not be applied blindly, but together with the sonority principle. So, the general rule is applied only when consonantal cluster violates the sonority principle we evoked in introduction.

Exception rules 4 and 5 are related to the particular status of Obstruent + Liquid + Glide sequence in French phonology, this consonant sequence being most often homosyllabic. Exception rule 5 presents an exception to Maximum onset principle and is largely motivated by the fact that we work with continuous speech and there is no knowledge about word boundaries integrated in our approach. In fact, consonant sequence Plosive + Liquid + Plosive is impossible word-internally in French, though it occurs in the corpus given the reduction phenomena typical in continuous informal speech.

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. We do not suggest that our solutions are the only ones particularly for syllables with complex structures as long as they are fairly uncommon in this specific corpus. Similarly, we added three specific rules to deal with phoneme sequences to which our rules do not apply. , that in some cases, rules are not relevant. Specifically for our corpus, the following phoneme sequences can not be separated: /pt/, /sk/ except when /pVsk/ and /fs/ as they correspond to frequent lexical units.

4. The LPL-Syllabeur Tool

4.1. Introduction

The program *LPL-Syllabeur*[1] is implemented in Java 1.6 and was tested under linux and windows®. This section

[1]The *Syllabeur* 'syllabifier' developped at LPL (Laboratoire Parole et Langage)

	Observed sequence	Segmentation rule	Examples (French)
1	VV	V.V	poëte: po.ɛt, il y a un: i.ɑ.œ̃, en haut: ɑ̃.o
2	VXV	V.XV	limité: li.mi.te, et donc on: e.dɔ̃.kɔ̃
3	VXXV	VX.XV	jardin: ʒaʀ.dɛ̃, comme ça: kom.sɑ, parce qu'il: pɑs.ki
4	VXXXV	VX.XXV	avec moi: ɑ.vek.mwa, cheval noir: sə.vɑl.nwɑʀ
5	VXXXXV	VX.XXXV	il se présentait: il.spʀe.zɑ̃.te
6	VXXXXXV	VXX.XXXV	alors je crois : ɑ.lorʒ.krwɑ

Table 1: General Rules

	Observed sequence	Segmentation rule	Examples
1	VXGV	V.XGV	baignoire: be.nwɑʀ, spéciaux: spe.sjo, tu vois: tu.vwɑ
2	VFLV	V.FLV	découvre: de.ku.vʀə,
3	VOLV	V.OLV	il trouve: i.tʀuv, mais de la: me.dlɑ
4	VFLGV	V.FLGV	effroyable: ef.ʀwɑ.jɑbl
5	VOLGV	V.OLGV	incroyable: ɛ̃.kʀwɑ.jɑbl
6	VOLOV	VOL.OV	connaître tu: ko.netʀ.ty, capable parce: ka.pabl.pɑs

Table 2: Exception Rules

describes the version 2.1 of the tool. *LPL-Syllabeur* is a jar file including the Java class sources, 3 description files, a documentation and 2 examples. The only accepted input/output file format is the Praat TextGrid format. This is a text file with a header and entries with 3 values: start time, end time and a label. We will call this *an object*. The object of the *LPL-Syllabeur* input file is the phoneme start time, the phoneme end time and the phoneme string label (usually one or two characters). The object of the *LPL-Syllabeur* output file is:

1. the syllable start time, which is the start time of its first phoneme,
2. the syllable end time, which is the end time of its last phoneme,
3. the syllable string label, which is the phoneme string labels concatenation.

This program is under GPL license and distributed freely.

4.2. How to use LPL-Syllabeur

A simple double-click will launch the program, with a GUI in English (Figure 1) or French. A first window asks the user to give:

- a configuration file name, containing the list of phonemes, their classes and all the rules,
- the phoneme input file name,
- the tier number, in case of the input praat file is of type *long* instead of *short*.

The second window shows all steps when the program is processing. Three output files are written in the same directory as the input one. Two files are statistics about n-grams phonemes and about rules used during the syllabification process. The last is the syllable file, with the extension name *syll-v2.1.textgrid*.

4.3. LPL-Syllabeur configuration file

We give here some details about the configuration file because this is a key part of the program. This is a simple ASCII text file that the user can change as needed. At first, the list of phonemes and the class symbol associated with each of the phonemes are described as, for example:

```
PHONCLASS e V
PHONCLASS p O
```

The couples phoneme/class are made of 3 columns: the first column is the key-word `PHONCLASS`, the second column is the phoneme symbol, the third column is the class symbol. The constraints on this definition are:

1. a vowel is mentioned with the class-symbol `V`,
2. a pause is mentioned with the class-symbol `#`,
3. if a phoneme contains the character =, it will be removed,
4. a class-symbol is only one character, excluding the character `X`.

The second part of the configuration file contains the rules. The first column is a keyword, the second column describes the classes between two vowels and the third column is the boundary location. The first column can be: `GENRULE`, `EXCRULE` or `OTHRULE`. In the third column, a 0 means the boundary is just after the first vowel, 1 means the boundary is one phoneme after the first vowel, etc. Here are some examples, corresponding to the rules described in this paper for spontaneous French:

```
GENRULE VXV 0
GENRULE VXXV 1
EXCRULE VFLV 0
EXCRULE VOLGV 0
```

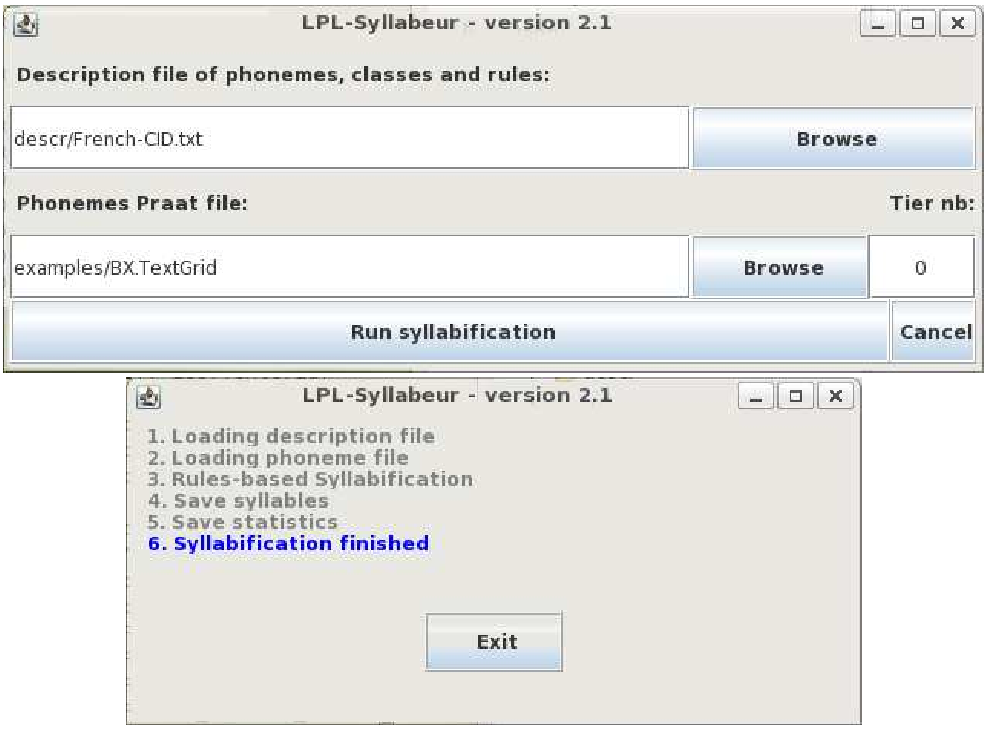

Figure 1: The LPL-Syllabeur GUI

Finally, to adapt the rules to specific situations that the rules failed to model, we introduced some phoneme sequences and the boundary definition. Specific rules contain only phonemes or the symbol `ANY` which means any phoneme. It consists of 6 columns: the first one is the keyword `OTHRULE`, the 4 following columns are a phoneme sequence where the boundary should be applied to the third one by the rules, the last column is the shift to apply to this boundary. In the following example: `OTHRULE ANY p s k -2`
the boundary will not be applied between /s/ and /k/ but before /p/ because in our corpus, very frequent words as *puisque* and *parce que* are phonetised /psk/.

The *LPL-Syllabeur* is distributed with the configuration file French-CID.txt corresponding to French phonemes encoded with SAMPA (Speech Assessment Methods Phonetic Alphabet), and the classes and rules described in this paper. Moreover, we included the configuration files French-limsi.txt, related to rules proposed in Table 1 of the paper (Adda-Decker et al., 2005) and French-C.Pallier.txt, related to rules proposed in (Pallier, 1999). The work is supported by the ANR research grant, *OTIM project*, ANR BLAN08-2_349062. The *LPL-Syllabeur-v2.1.jar* is downloadable at: http://lpl-aix.fr/˜otim.

5. Evaluation

5.1. Test corpus description

The test corpus is 1.6% of the CID. It is about 7 minutes of a dialogue, which represents 653 words for speaker 1 and 1,238 words for speaker 2. This test corpus contains 2,068 syllables. Our test corpus is available in the jar file of the *LPL-Syllabeur*.

The test corpus was manually segmented by two experts, and 23 boundary mismatches are observed (this means 46 different syllables). A boundary mismatch means that one expert does not propose the same segmentation between two syllables as the other expert. So the syllable agreement rate is 97.77% (46 over 2068). Of course, mismatches increase depending on the number of consonants between two vowels. Details about these mismatches are reported below:

- 5 mismatches with VXV, over 1165 cases, 99.57% agreement,
- 12 mismatches with VXXV+exceptions over 435 cases, 97.24% agreement,
- 5 mismatches with VXXXV+exceptions over 43 cases, 88.37% agreement,
- 1 mismatch with VXXXXV over 3 cases, 66.67% agreement.

Other segmentations are obvious and it is unsurprising that the two experts have an agreement rate of 100%: a vowel followed by a vowel, a vowel followed by a pause, a pause followed by a vowel or a pause followed by another pause (with consonants between the two, like *mh* to say yes). The table 3 show statistics about the rules used to syllabify the test corpus.

5.2. Syllabification performances

Other systems were adapted and evaluated as we want to have an idea about their performances on our specific corpus. " To adapt these systems, we made the following:

	syllabation.awk (Pallier, 1999)	*graphon+* (Adda-Decker et al., 2005)	*syllabify2.praat* (Goldman, 2007)	*LPL-Syllabeur*
Expert 1	74	80	67	43
	7.16%	7.74%	6.48%	**4.16%**
Expert 2	84	85	75	53
	8.12%	8.22%	7.25%	**5.13%**

Table 4: Number of boundary mismatches and syllable difference rate (percentage)

Number	Rule		Number	Rule
145	V#,#V,##			
276	VV			
1165	VXV			
435	VXXV	including	54	VXGV
			17	VFLV
			73	VOLV
43	VXXXV	including	0	VFLGV
			4	VOLGV
			4	VOLOV
3	VXXXXV			

Table 3: Statistics about rules in the test corpus

- *syllabation.awk* from (Pallier, 1999) is freely available, so we adapted its phoneme encoding (SAMPA) and also converted our data format,
- *graphon+* is not free, so we made a configuration file for our syllabification program, by using the rules described in Table 1 of the paper (Adda-Decker et al., 2005),
- *syllabify2.praat* from (Goldman, 2007) is free and uses praat, so we just adapted this tool to our phoneme encoding.

All results are reported in Table 4. Results show improvements when using our system. In this case, an improvement means that the automatic syllabification is nearest to the manual one. Finally, the rules developed here seems to be well-adapted to the syllable boundary detection of this difficult corpus.

5.3. Examples

In Table 5 we present some examples of syllabification extracted from the automatic system. These examples are compared with the result of the two experts' syllabification. In the first example, the automatic system and experts are in agreement. In the second one, both experts propose a different syllabification from the system output. In this second example, we observe that experts are influenced by lexical boundaries. And in the third example, the second expert disagrees with both automatic system and first expert. Most of the differences between automatic and manual syllabification concern word juncture. Human experts tend to be influenced by word boundaries not only when syllable boundaries are unclear, but even when they concern unambiguous boundaries.

Here are some examples of ambiguous boundaries, for which experts are influenced by lexical boundaries:

- Transcription: **(5)** *sur Aix mais* 'in Aix but'
  - Syllables expert1 and expert2: ʀeks . me
  - Syllables auto: ʀek . sme
- Transcription: **(6)** *glaces comme ça* 'ice like this'
  - Syllables expert1 and expert2: glɑs . kom
  - Syllables auto: glɑ . skom
- Transcription: **(7)** *créatif ouais #* 'creative yeah'
  - Syllables expert1 and expert2: tif . we
  - Syllables auto: ti . fwe

Here are some examples of unambiguous boundaries for experts. They are influenced by lexical boundaries:

- Transcription: **(8)** *vous offre le* 'offers you the'
  - Syllables expert1 and expert2: zof . lə
  - Syllables auto: zo . flə
- Transcription: **(9)** *comme une folle* 'like a mad girl'
  - Syllables expert1 and expert2: kom . yn
  - Syllables auto: ko . myn

Moreover, in most of the cases, mismatches between the automatic syllabification and the experts' syllabification concern ambiguous boundaries for which experts propose variable syllabification as it has been shown in some perception experiments (Content et al., 2001). Here are some examples of ambiguous boundaries, for which experts do not agree:

- Transcription: **(10)** *retrouver les* 'to find the'
  - Syllables expert1: ɑʀt . ʀu
  - Syllables expert2: ɑ . ʀtʀu
  - Syllables auto: ɑʀ . tʀu
- Transcription: **(11)** *va se faire* 'will be'
  - Syllables expert1 and auto: vɑs . feʀ
  - Syllables expert2: vɑ . sfeʀ
- Transcription: **(12)** *pas le truc* 'not the thing'
  - Syllables expert1 and auto: pɑl . tʀyk
  - Syllables expert2: pɑ . ltʀyk

TOE	et donc on mange sur la [baignoire, bainoire] donc c’est c’est ça
Phonemes	e d ɔ̃ k ɔ̃ m ɑ̃ ʒ s y ʀ l ɑ b e n w ɑ ʀ d ɔ̃ k s e s e s ɑ
Classes	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 F V
Syllables (Auto & Experts)	e . dɔ̃ . kɔ̃ . mɑ̃ʒ . syʀ . lɑ . be . nwɑʀ . d ɔ̃ k . se . se . sɑ
TOE	non dans les parcs c’est un peu limité
Phonemes	n ɔ̃ d ɑ̃ l e p ɑ ʀ k s e t œ̃ p ə l i m i t e
Classes	N V O V L V O V L O F V O V O V L V N V O V
Syllables Auto	nɔ̃ . dɑ̃ . le . pɑʀ . kse . tœ̃ . pə . li . mi . te
Syllables Expert1 & Expert2	nɔ̃ . dɑ̃ . le . pɑʀk . se . tœ̃ . pə . li . mi . te
TOE	il expliquait pas vraiment [c(e), ss] qu’i(l) y avait d(e)dans
Phonemes	i l e k s p l i k e p ɑ v ɾ e m ɑ̃ s k i j ɑ v e d ɑ̃
Classes	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
Syllables Auto	i . lek . spli . ke . pɑ . vɾe . mɑ̃ . ski . jɑ . ve . dɑ̃
Syllables Expert1	i . lek . spli . ke . pɑ . vɾe . mɑ̃ . ski . jɑ . ve . dɑ̃
Syllables Expert2	i . leks . pli . ke . pɑ . vɾe . mɑ̃ . ski . jɑ . ve . dɑ̃

Table 5: Some syllabification examples

This group of examples shows that when there are sequences of more than 2 consonants (not quite frequent in French, though often resulting from reduction phenomena), one of the experts pays more attention to the lexical material, while the second expert (and our algorithm) favours a more balanced structure and respects the sonority principle. Syllable boundaries extracted from automatic system do not depend on lexical boundaries as long as the system did not have this information. The input to the system consists only of a string of phonemes. In this way we do not consider that expert syllabification is a reference of what should be a correct syllabification. Expert’s syllabification is the result of various informations (phoneme string, orthographic representation, lexical boundaries, discourse organisation and possible pauses, etc.) and a result of a complex linguistic and cognitive task (Treiman and Danis, 1988). This is not exactly the aim of the tool we present: we expect to provide a syllabified corpus in a way to allow phonetic, prosodic or various linguistic analyses. Lexical, discourse and other linguistic information are available in the CID corpus. Consequently, we have the possibility to evaluate the role of lexical and syllabic (according to our rules) boundaries independently.

5.4. Statistics and analysis

In total, 1,600 syllable boundaries are currently concerned with the rules: VXV, VXXV, VXGV, VOLV and VFLV. For these rules, there is no ambiguity (see details in table 6). This means that the result of syllabification would be completely compliant with usual phonological rules as well as to human performance.

The main syllabification problem particularly concerns the occurrence of two vowels separated by more than two consonants. These occurrences are rare and the most frequent one is VXXXV (in which the cases with plosives and fricatives in second consonant position are excluded). The proposition we made is to put the syllable boundary after the first consonant. We hypothesize that it would not be convenient for some occurrences since they appear in word junctures. The possible wrong syllabifications will be corrected after the expert analysis.

	Expert 1	Expert 2	Total nb
VXV	5	4	1165
	0.43%	0.34%	
VXXV+Exceptions	26	32	435
	5.98%	7.36%	
VXXXV+Exceptions	11	15	43
	25.59%	34.88%	
VXXXXV	1	2	3
	33.33%	66.67%	

Table 6: Mismatch statements of the *LPL-Syllabeur*

6. Conclusion

The work presented in this paper is a rule-based phoneme to syllable segmentation system. We showed that rules we propose are particularly well-adapted to the syllabification of a spontaneous French corpus in a friendly dialogue context. Compared with existing systems, the advantages of the *LPL-Syllabeur* are that (1) it is made with a small number of simple rules, (2) the tool uses an object-oriented language (3) it is under GPL license and (4) it is very easy to adapt to a specific corpus by adding or modifying rules, phoneme encoding or phoneme classes, by the use of a new configuration file. Automatic system output and expert’s syllabification are in agreement for most of syllable boundaries in our corpus. The differences appear (1) when an intervocalic sequence is constituted by more than two consonants (these syllable boundaries may appear ambiguous, even for experts), or (2) when there is an “incongruency” between syllable and lexical boundaries (in some cases, experts tend to favour lexical boundaries). Consequently, the output of the *LPL-Syllabeur* gives a regular and faithfull syllabification of a spontaneous speech corpus. Ambiguous boundaries should be treated in specific works as well as the influence of lexical boundaries, which have to be analysed with the speech signal. Syllabification is an essential component of many speech and language processing systems, and this tool might be very useful to researchers working on syllabification on various languages.

7. References

M. Adda-Decker, P. Boula de Mareüil, G. Adda, and L. Lamel. 2005. Investigating syllabic structures and their variation in spontaneous french. *Speech Communication*, 46:119–139.

S. Bartlett, G. Kondrak, and C. Cherry. 2008. Automatic syllabification with structured SVMs for letter-to-phoneme conversion. In *Proceedings of Human Language Technologies: The 2008 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 568–576, Columbus, Ohio, USA. Association for Computational Linguistics.

S. Bartlett, G. Kondrak, and C. Cherry. 2009. On the syllabification of phonemes. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 308–316, Boulder, Colorado, USA. Association for Computational Linguistics.

R. Bertrand, P. Blache, R. Espesser, G. Ferr, C. Meunier, B. Priego-Valverde, and S. Rauzy. 2008. Le CID - corpus of interactional data. *Traitement Automatique des Langues*, 49(3):105–134.

P. Blache, R. Bertrand, and G. Ferr. 2008. Creating and exploiting multimodal annotated corpora. In *Language Resource and Evaluation Conference*, pages 38–53, Marrakech (Marroco). Multimodal Corpora, Kipp M. (Ed.).

C. Blanche-Benveniste. 1990. *Le franais parl, tudes grammaticales*.

P. Boersma and D. Weenink. 2009. Praat: doing phonetics by computer, http://www.praat.org.

P. Boula de Mareüil. 1997. Etude linguistique applique la synthse de la parole partir du texte.

A. Content, R-K. Kearns, and U-H. Frauenfelder. 2001. Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language*, 45:177–199.

D. Crystal. 2003. *A Dictionary of Linguistics and Phonetics (5th edn.)*. Oxford: Blackwell Publishing.

J-P. Goldman. 2007. http://latlcui.unige.ch/phonetique/.

J. Goslin and U.H. Frauenfelder. 2001. A comparison of theoretical and human syllabification. *Language and Speech*, 44(4):409–436.

D. Kahn. 1976. Syllable-based generalizations in english phonology. *Indiana University Linguistics Club*.

Y. Marchand and R.I. Damper. 2007. Can syllabification improve pronunciation by analogy of english? *Natural Language Engineering*, 13(1):1–24.

Y. Marchand, C.R. Adsett, and R.I. Damper. 2009. Automatic syllabification in english: A comparison of different algorithms. *Language and Speech*, 52(1):1–27.

2004. Lexique 2: A new french lexical database. *Behavior Research Methods Instruments & Computers*, 36(3):516–524.

C. Pallier. 1999. Syllabation des reprsentations phontiques de brulex et de lexique, http://www.pallier.org.

E. Pulgram. 1970. *Syllable, Word, Nexus, Cursus.* Mouton & Co, La Haye, Pays-Bas.

A.M. Sandoval, D.T. Toledano, R. de la Torre, M. Garrote, and J.M. Guirao. 2008. Developing a phonemic and syllabic frequency inventory for spontaneous spoken castilian spanish and their comparison to text-based inventories. In *Language Resource and Evaluation Conference*, pages 1097–1100, Marrakech (Marroco).

R. Treiman and C. Danis. 1988. Syllabification of intervocalic consonants. *Journal of Memory and Language*, 27:87–104.