

# Phonetic transcription & syllabification protocol for the project

## *SyllabO+*

Version August 4th 2016



LABORATOIRE DES NEUROSCIENCES  
DE LA PAROLE ET DE L'AUDITION

---

SPEECH AND HEARING  
NEUROSCIENCE LABORATORY



UNIVERSITÉ  
LAVAL

## ***1. General principles***

- 1.1. All phonetic transcriptions were executed by research assistants trained in linguistics and phonetics, under the supervision of Pascale Tremblay, Ph. D., and Johanna-Pascale Roy, Ph. D. The transcriptions were made from the audio recordings that constitute our corpus.
- 1.2. The transcriptions are an accurate and non-normative reproduction of the speaker's vocabulary and syntax.
- 1.3. Prosody is not transcribed and neither are silences, laughs, onomatopoeia and other prosodic markers and punctuators. However, depending on the context, words with a specific function and meaning are transcribed. Examples: “*ah*”, “*euh*”, “*t’sais*”, “*ouais*”, “*pis*”, “*eille*”, “*oh*”, “*genre*”, “*comme*”, “*style*”, “*ok*”, etc.
- 1.4. Non-speech sounds (breathing, coughing or other non-speech elements) are not transcribed, as well as hesitations, stuttering and other background noises made by the speaker or other people present during the recordings.
- 1.5. When transcriptions were obtained from a recording where the speaker was discussing in interaction with one or several other people, the distinct speech turns are kept in the transcription. Thus, only one speaker is transcribed, but his speech is not transcribed in one large unit, but rather in sections corresponding to his speech turns.
- 1.6. All files are transcribed in International Phonetic Alphabet (IPA). *See next page for IPA chart*<sup>1</sup>.
- 1.7. All transcribed words are segmented into individual syllables.

---

<sup>1</sup> IPA Chart, <http://www.internationalphoneticassociation.org/content/ipa-chart>, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2015 International Phonetic Association



## 2. *Phonetic Transcription*

2.1. Phonetic transcriptions are an exact, but broad reproduction of the audio recording. Regional, generational, or other variations are ignored.

2.1.1. Affrication: non-distinctive affrication is not transcribed. (*Ex: The word “tu” is transcribed [ty] even if pronounced [tsy].*) because affrication of [t] and [d] in front of [i] and [u] is systematic in Quebec French and is not distinctive (Dumas, 1987).

2.1.2. Vowel reduction: vowel reduction (vowels [i], [y], and [u] that become [ɪ], [ʏ], and [ʊ] in a closed syllable) is not transcribed (*Ex: The word “vite” is transcribed [vit] even if pronounced [vɪt].*) because reduction of vowel [i], [y], and [u] is systematic in Quebec French and is not distinctive (Dumas, 1987).

2.1.3. Vowel variation is transcribed according to speaker’s pronunciation.

- Examples:
  - [a] vs. [ɑ] (*Ex: The word “chat” is transcribed [ʃa] or [ʃɑ] depending on the speaker’s pronunciation.*) (Note that the variant [ɔ] is transcribed [ɑ] in all cases where the phoneme /a/ or /ɑ/ is pronounced [ɔ]);
  - [ɛ] vs. [e] (*Ex: The word “avait” is transcribed [avɛ] or [ave] depending on the speaker’s pronunciation.*);
  - [ɛ] vs. [ɛ̃] (*Ex: The word “maison” is transcribed [mɛzɔ̃] or [mɛ̃zɔ̃] depending on the speaker’s pronunciation.*);
  - etc.

2.1.4. For phonemes that have different possible allophonic variants, one allophone was selected for transcription:

- [ə] is selected rather than [œ] to transcribe the phoneme /ə/ (*Ex: The word “je” is transcribed [ʒə], whether the speaker’s pronunciation is [ʒə] or [ʒœ].*);

- [r] is selected rather than [ʀ] or [ʁ] to transcribe the phoneme /r/ (Ex: The word “rue” is transcribed [ry], whether the speaker’s pronunciation is [ry] or [ʀy] or [ʁy].);
- [ã] is selected rather than [ā] to transcribe the phoneme /ã/ (Ex: The word “lent” is transcribed [lã], whether the speaker’s pronunciation is [lã] or [lā].);
- [ẽ] is selected rather than [ē] to transcribe the phoneme /ẽ/ (Ex: The word “main” is transcribed [mẽ], whether the speaker’s pronunciation is [mẽ] or [mē].).

2.1.5. Phonemes that can be transcribed by equivalent characters – *i.e.* identical from the IPA’s standpoint, but different in terms of computer characters – are systematically transcribed by only one same character.

- Phoneme /g/ is transcribed [g] and not [g].

2.2. Elided segments are not reconstructed in the phonetic transcription.

- Example:
  - The word “ministre” is transcribed [minis] if that was the pronunciation; phonemes /t/ and /r/ are not transcribed if they have been elided.

2.3. Sentences separated by a pause, but semantically connected, are separated by a vertical line (|) in the original document. *Note that these vertical lines do not appear in the tables of the database.*

2.4. Sentences that have no logical relation (topic change) or that are part of different speech turns during an interaction are separated in distinct “speech turns” in the original document to indicate that there is no link between the last syllable of a section and the first syllable of the next section. *Note that these “speech turns” do not appear in the syllable tables of the database since each speech turn has been separated in individual syllables or groups of syllables.*

2.5. Unintelligible words are transcribed as a hash symbol (#) in the phonetic transcription<sup>2</sup>.

However, if individual sounds of the word are intelligible, the sound sequence is transcribed, as long as individual syllables are recognized, even if no complete word can be recognized from these syllables. If necessary, syllabification is completed as usual on this “meaningless word”. If too many sounds are unintelligible (incomplete syllable), the whole sound sequence (syllable(s)) should be indicated as unintelligible (#).

2.6. Segment lengthening and non-distinctive diphthongs are not transcribed. *Note that distinctive long segments are transcribed (see example), but do not appear in the syllable tables of the database.*

- Example:

- “maître” or “mètre” and “mettre” → [mɛ:tr] or [mɛ:tr] and [mɛtr]

2.7. English IPA is used for English names, as well as for words produced with an English pronunciation<sup>3</sup>.

---

<sup>2</sup> Across the corpus (184 recordings), 76 unintelligible segments were noted (#), representing 0.025023% of the 303,636 syllables.

<sup>3</sup> Note: the proportion of English-pronounced syllables (containing one or more typically English sound) in the corpus is 458 out of 303 636 syllables (0,15%). It is possible to filter out these syllables (or phones) in the database table by excluding any that contain the English-specific consonants and vowels used in the transcriptions, which are: [ɹ ð h] and [ʌ ɒ ɜ æ ɪ ʏ ʊ].

### 3. Syllabification

#### 3.1. General rules:

3.1.1. Syllabification was performed according to the existing structure of French syllables, prioritizing consonants at onset rather than at coda (Brousseau & Nikiema, 2001; Paradis, 1993);

- Example:
  - “catalogue” → [ka ta lɔg] (*and not \*[kat al ɔg]*)

3.1.2. Within a word, the least marked syllables (CV, CVC and CCV) are prioritized.

- Examples:
  - “adversaire” → [ad vɛr sɛr];
  - “budget” → [byd ʒɛ];
  - “athlétique” → [at le tik].

#### 3.2. Schwa elision in oral speech:

3.2.1. Within a phrase, the single phoneme remaining after the elision of the schwa [ə] is attached to the next syllable. In this case, branching onsets are prioritized.

- Examples:
  - “ça **me** fait rire” → [sa **m**fɛ rir];
  - “tu **te** souviens” → [ty **tsu** vjɛ̃];

3.2.2. Within a word, the single phoneme remaining after the elision of the schwa [ə] is attached to the preceding syllable. In this case, CVC syllables are prioritized.

- Examples:
  - “samedi” → [sam **d**i];
  - “l’avenir” → [lav **n**ir];

- “entretenir” → [ã trət nir].

3.2.3. These distinctions account for certain distinctive sequence pairs:

- Examples:
  - “parce que” [pars kə ] versus “par ce que” [par skə];
  - “samedi” [sam di] versus “ça me dit” [sa mdi].

3.2.4. Syllables forming lexicalized words in Quebec French remain as are:

- Examples:
  - “fait que” [fæk] or [fak];
  - “rien que” [ɛ̃k] or [jɛ̃k].

### 3.3. Elision of several parts of a word or phrase in oral speech:

3.3.1. When several phonemes are elided from a sequence, the syllable is formed according to the resulting sounds.

- Examples:
  - “sur le pouce” [sylpus] → [syl pus];
  - “où est-ce que” [uskə] → [us kə];
  - “que c’est un” [kstɛ̃] → [kstɛ̃].

### 3.4. Lengthening:

3.4.1. Lengthened segments (at the end of a syllable) are doubl/ed during the process of syllabification, either to be part of two syllables or to be part of one syllable and form another syllable on its own.

- Examples:
  - “extrêmement” [ɛkstrɛm:ã] → [ɛk strɛm mã];



- “nécessairement” [nes:ɛrmã] → [nɛs sɛr mã];
- “ça a participé” [sa:partisipe] → [sa a par ti si pe];
- “on a été” [ɔ̃ne:te] → [ɔ̃ ne e te].

### 3.5. Words with “x” ([ks] or [gz]):

3.5.1. For sequences including the sound “x”, whether voiced [gz] or unvoiced [ks], the onset consonants are prioritized (Brousseau & Nikiema, 2001; Paradis, 1993). Thus:

3.5.1.1. When onset is already filled, the /s/ is part of the preceding syllable:

- Examples:
  - “exclu” → [ɛks kly];
  - “exquis” → [ɛks ki];
  - “extrêmement” → [ɛks tɛm mã];
  - “exploit” → [ɛks plwa];
  - etc.

3.5.1.2. When the onset is empty, it is then filled by the /s/ (or the /z/):

- Examples:
  - “décomplexé” → [de kɔ̃ plɛk sɛ] (*and not \*[de kɔ̃ plɛks e]*);
  - “klaxonnant” → [klak sɔ̃ nã] (*and not \*[klaks ɔ̃ nã]*);
  - “réflexion” → [re flɛk sjɔ̃] (*and not \*[re flɛks jɔ̃]*);
  - “exemple” → [ɛg zãpl] (*and not \*[ɛgz ãpl]*).
  - etc.

### 3.6. Rules specific to French liaisons:

3.6.1. “Illegal” French syllables are avoided.

- Example:

- “mon ami” → [mɔ̃ na mi] (*and not \*[mɔ̃n a mi]*).

3.6.2. In case of a non-epenthesis (no added sound), the lexicon is prioritized.

- Examples:

- “quinze ans” → [kɛ̃z ɑ̃] (*and not \*[kɛ̃z zɑ̃] or \*[kɛ̃ zɑ̃]*);
- “par amour” → [paʁ amur] (*and not \*[pa ra mur]*);
- “entre autres” → [ɑ̃tr otr] (*and not \*[ɑ̃t rotr]*);
- “bon an” → [bɔ̃ nɑ̃] (epenthesis) or [bɔn ɑ̃] (non-epenthesis), based on the actual pronunciation (*and not \*[bɔ nɑ̃]*).
- “quand” and “quant” :
  - “quand on prend...” → [kɑ̃ tɔ̃ prɑ̃] (epenthesis)
  - “quant à lui...” → [kɑ̃t a lui] (non-epenthesis)

3.6.3. In case of an epenthesis (added sound), the empty onset is filled and the liaison becomes part of the following word.

- Examples:

- “les amis” → [le za mi] (*and not \*[lez ami]*);
- “au cœur de la ville” → [o kœʁ də la vil] (*added phoneme*);
- “ensuite de ça” → [ɑ̃sɥit də tsɑ] (*added phoneme*);
- “un œuf” → [œ̃ nœf].
- “bon an” → [bɔ̃ nɑ̃] (epenthesis) or [bɔn ɑ̃] (non-epenthesis), based on the actual pronunciation (*and not \*[bɔ nɑ̃]*).

## References

Brousseau, A.M., & Nikiema, E. (2001). *Phonologie et morphologie du français*: Fides.

Dumas, D. (1987). *Nos façons de parler: les prononciations en français québécois*. Sillery: Presses de l'Université du Québec.

Paradis, C. (1993). Phonologie générative multilinéaire. In J.-L. Nespoulous (Ed.), *Tendances actuelles en linguistique générale* (pp. 11-47). Neuchâtel: Delachaux et Niestlé.