

User manual for *SyllabO+*

Version of January 20, 2018



LABORATOIRE DES NEUROSCIENCES
DE LA PAROLE ET DE L'AUDITION

SPEECH AND HEARING
NEUROSCIENCE LABORATORY

User manual

Before downloading the full database, it is recommended to have an idea of what data is needed for your project (given its large size). The databases contain a great amount of information, and it is possible that only a portion of it will be useful for a particular purpose.

Instructions

1. Go to <http://syllabo.speechneurolab.ca/> and after filling in a few questions, you will access the **SyllabO+ databases**. All the information will remain confidential and will only be used to compute statistics and report to funding agencies.

Accéder à / Go to SyllabO+

2. You may then **choose** to either **download a complete database** (information extracted from all the audio recordings) – whether **phones, syllables, words, or lemmas**– **or browse and download a specific database**. The complete database contains statistics about the use of units across communication contexts, age and gender. It represents the average use of oral language in native adult speakers of Quebec French, in terms of units (phones/syllables/words/lemmas) and units sequences.

The example below shows the complete syllable database.

Syllable database

When you download the entire syllable database, three files are included in a .ZIP folder: the single syllables, the sequences of two syllables (pairs), and sequences of three syllables (triads).

Alternative downloads are also available below in the form of .XLSX files (with special IPA characters already embedded) and .CSV files transcribed in **SAMPA** – Speech Assessment Methods Phonetic ALphabet (without any special characters).

Download the SYLLABLE database

Alternatively, you can **browse and download a specific database**, by clicking on this button:

To search online, click here

The database will contain statistics computed from a subset of the audio recordings, by using the following parameters, individually or combined: context of communication (formal, informal), age (range), and sex of the speakers. This means that you could obtain the statistics, for example, of syllable use for only the young adults, or for all women in formal contexts, to name a few of the many possibilities. Online viewing of the results is also available.

The image below shows the form to download or browse the desired specific database.

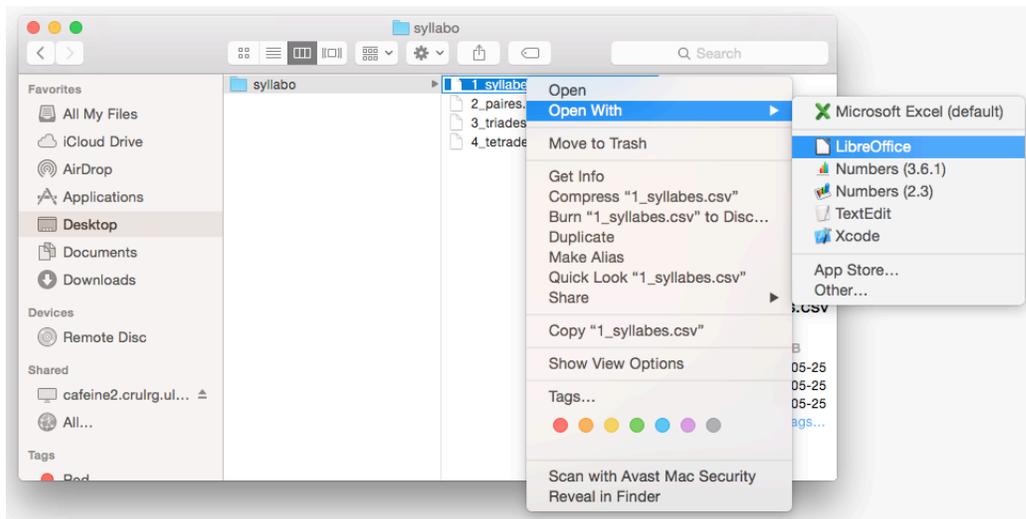
The image shows a search form with the following elements:

- Type***: A dropdown menu with "Phones - Phones" selected.
- Contexte d'élocution - Communication context**: A dropdown menu with "Tous - All" selected.
- Genre des locuteurs - Speaker gender**: A dropdown menu with "Tous - All" selected.
- Résultats par page - Results per page***: A dropdown menu with "20 résultats - 20 results" selected.
- Âge minimum - Minimum age**: A dropdown menu with "-----" selected.
- Âge maximum - Maximum age**: A dropdown menu with "-----" selected.
- Voir en ligne - Browse online**: A blue button.
- Exporter en CSV* - Export as CSV***: A blue button.

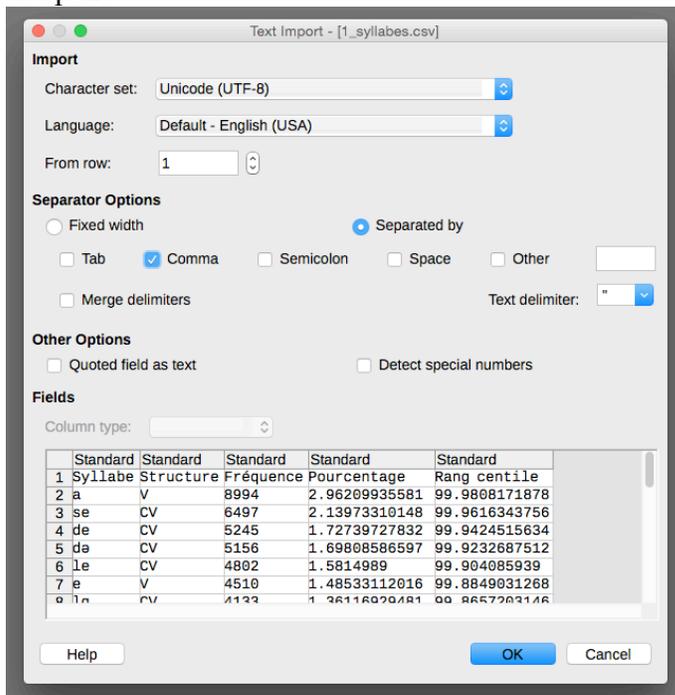
3. If you have **downloaded a complete database**, you will obtain a few different files compressed in ZIP format for convenience. For the syllables, there will be one file for the individual syllables and their statistics, one for the pairs of syllables and their statistics, and one for the groups of 3 syllables (triads) and their statistics. For the phones, you will obtain two files, one for the individual phones and their statistics, and one for the pairs of phones and their statistics.

For the words or lemmas, there will be one file for the individual words/lemmas and their statistics, one for the pairs of words/lemmas and their statistics, and one for the groups of 3 words/lemmas (triads) and their statistics.

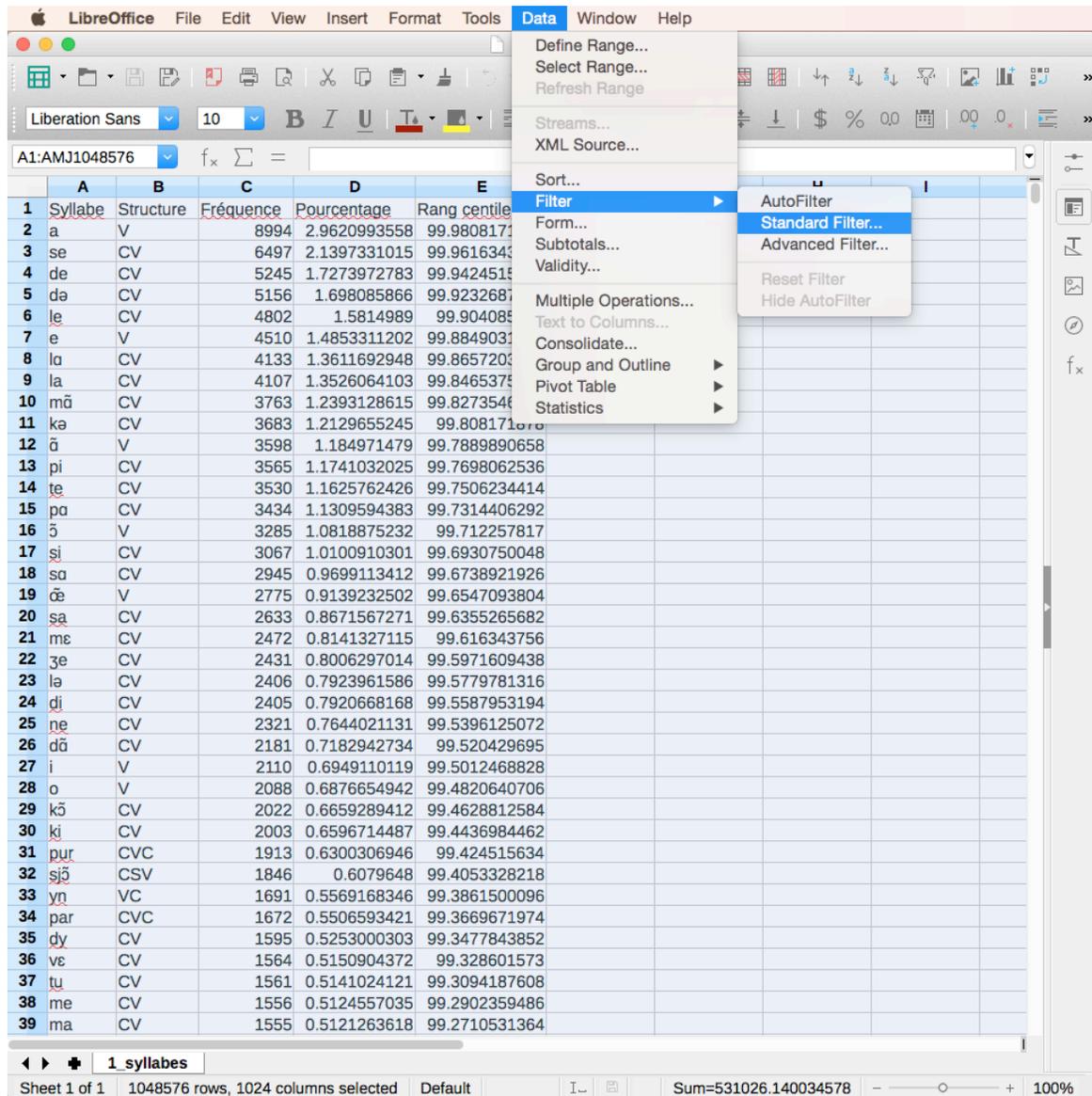
4. If you **downloaded a specific database**, you will only obtain the selected file (single units, pairs of units, or triads).
5. All files are in **CSV format** (Comma Separated Values), which is a simple way of storing tabular data in plain text files. In these files, units are transcribed in International Phonetic Alphabet (IPA) (for phones and syllables only, non applicable for words and lemmas). Alternative downloads are also available in the form of .XLSX files (with special IPA characters already embedded) and .CSV files transcribed in SAMPA – Speech Assessment Methods Phonetic Alphabet (without any special characters). However, these formats are only available to download the entire database.
6. To open the CSV (IPA) files we recommend using the software **LibreOffice**. We recommend it because it handles all phonetic (IPA) characters correctly. Important: Microsoft Excel does not handle these special characters properly, (unless using the XSLX files). This is why we provide instructions to guide you through the process of having the characters recognized, and this is also why we provide alternative downloads. Note: LibreOffice is a free and open source software available for download at <http://www.libreoffice.org/> .



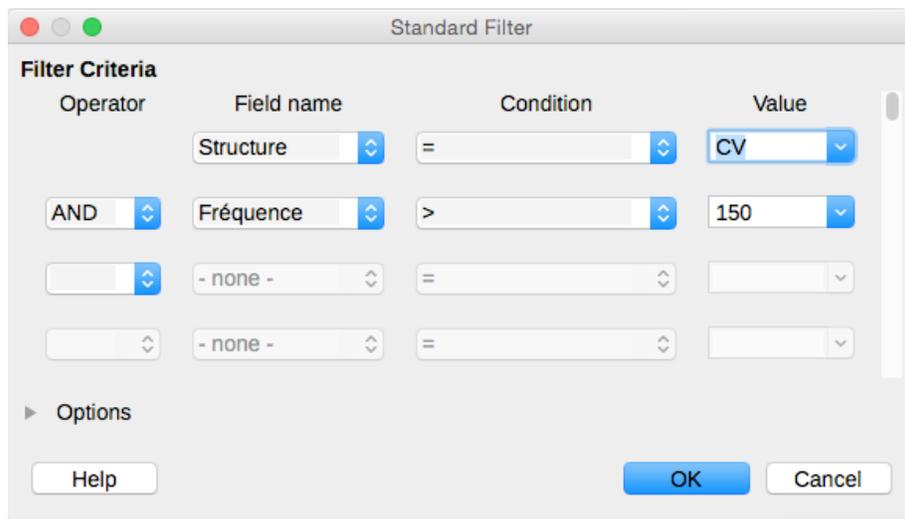
7. Select **Unicode (UTF-8)** as the character set. Select **Comma** as the separator option. Click **OK** to open.



8. Once the file is opened, you can select only the columns relevant to your project (*Hide* or *Delete* unwanted columns).
9. Filter the data according to your needs: select all data, go to the menu, and click on **Data – Filter – Standard Filter**.



10. You may then use any **filter** you wish to obtain the desired information. Select which column is to be filtered, according to which parameter. For example:



11. You obtain your **results!**

Example of stimuli creation

Here is an example of how specific stimuli can be created from *SyllabO+*. Using the table containing groups of 3 syllables, these specific columns were selected:

- Group of 3 syllables (IPA transcription)
- Structure 1st syllable
- Structure 2nd syllable
- Structure 3rd syllable
- Percentile of score 1st syllable
- Percentile of score 2nd syllable
- Percentile of score 3rd syllable

The filter “ = CV” was applied to all “structure” columns and the filter “ >= 95” was applied to all “percentile of score” columns. The resulting table (*Table 1*) is thus a selection of syllable triads, where each syllable has a simple “consonant-vowel” structure and has a high frequency since it is found

between the 95th and 100th percentile. From this table, the groups of 3 syllables can be easily used as stimuli in an experimental study.

Table 1. Stimuli: triads of simple-structure (CV), high-frequency syllables

Triad	Structure 1	Percentile 1	Structure 2	Percentile 2	Structure 3	Percentile 3
kə mǎ se	CV	99.0216765778	CV	99.8273546902	CV	99.9616343756
vu za ve	CV	98.849031268	CV	98.6188375216	CV	98.273546902
mǎ dɔ ne	CV	99.8273546902	CV	97.602148475	CV	99.5396125072
le zǎ fǎ	CV	99.904085939	CV	98.4270093996	CV	97.0650297334
dɪ fe rǎ	CV	99.5587953194	CV	96.451179743	CV	98.0625359678
...

That's it! Hope you enjoy it, and stay tuned for future developments (easier online searches)! Please do not forget to cite our article in all your publications and reports derived from the use of *SyllabO+* (Bédard et al., 2016) :

Bédard, P., Audet, A.-M., Drouin, P., Roy, J. P., Rivard, J., & Tremblay, P. (2016). *SyllabO+*: A new tool to study sublexical phenomena in spoken Quebec French. *Behavioral Research Methods*.